2015

Characterizing the Temporal Dynamics in Functional Connectivity Measured with fMRI

Joshua Swearingen
Medical University of South Carolina

Characterizing the temporal dynamics in functional connectivity measured with fMRI.

Joshua Swearingen

A dissertation submitted to the faculty of the Medical University of South Carolina
in partial fulfillment of the requirement for the degree of Doctor of Philosophy
in the College of Graduate Studies.

Department of Neuroscience

2015

Approved by:

Jane Joseph Chair

Judson Chandler

Brian Dean

Andrew Lawson

Thomas Naselaris

TABLE OF CONTENTS

List of Figures

Acknowledgments

“From all my teachers have I gotten understanding.”

– Ben Zoma, *Ethics of the Fathers(second century AD)*

Abstract

There is a considerable and growing interest in the organization and development of neural function at the scale of the entire brain, particularly from activity observed in fluctuations in the blood oxygen level dependent (BOLD) signal in functional magnetic resonance imaging (fMRI). This activity reliably organizes into sub-networks, whether obtained from subjects at rest or from those involved in various perceptual or cognitive tasks. While these networks have often been identified and studied, the precise dynamics involved in their interactions as well as the relationship between organization found in both resting and task-based activity are not well understood.

Here I report not only the traditional functional organization of the resting brain as observed through inter-regional correlations, but how this organization changes over time. The dynamics of the brain during rest are not stationary as typically assumed, but vary as different sub-networks co-activate. To produce this more advanced model I apply a new method from the machine learning literature that uses spectral learning to estimate the latent dynamics of brain activity driving these changes. When comparing this model to one from subjects passively viewing faces and other objects, I find distinct changes in a sub-network containing regions of the brain involved in object recognition.

Additionally, the importance of fMRI signal variability as an independent within-subject measure has recently gained attention. Looking at those same subjects

at rest and passively viewing images, I introduce a simple measure that finds a differential network that may expose the sources of variability driving these dynamics. This network appears to be largely absent in anesthetized primates, and is disrupted in a cohort of cocaine users viewing drug related paraphernalia.

These results together suggest a picture of a dynamic brain, with multiple interacting subsystems that are not discrete isolated elements but often overlap. These superpositions of activity may give insights into the dynamics seen during various attentional tasks, where smaller parts of these networks tend to increase in activity. Finally, there may be an independent level of organization that is coordinating this dynamic activity as seen through the variability of these systems over time. This novel network both increases in magnitude and is predictive of age, and discriminates between faces and other objects. These results help further our understanding of network organization in the brain.

Background

1.1 Introduction

Understanding the interactions of neurobiological substrates that support human cognition and perception has been a fundamental neuroscience research question for decades. Techniques such as whole-brain functional magnetic resonance imaging provide an opportunity to study large scale organization and function of the human brain. However, fMRI studies generate large volumes of data which presents challenges for traditional statistical methods. One promising approach to address this difficulty comes from the study of complex systems and networks, where structure and function arise from the many interacting elements of the system. Literature suggests that the brain exhibits network structure at possibly every scale, from complex chains of molecular interactions, to populations of cells encoding stimuli, and ultimately to graph theoretic descriptions of the anatomical connectivity providing the scaffold for all activity. Particular interest has been given to those functional networks derived from the fMRI signal (Bullmore & Sporns, 2009) in part because of their ability to cover large spatial scales and non invasive nature of the imaging technique. Many of the networks discovered are present during spontaneous activity in resting states (Raichle et al., 2001), can show changes across development (Fair et al., 2009), and these resting state networks may account for the majority of brain activity ultimately utilized for various cognitive demands (Raichle, 2009). In this work I intend to better quantify this spontaneous activity, first by modeling

the temporal dynamics of information shared between brain regions, investigating intrinsic maps of functional activity and how this activity segregates the brain into multiple sub-networks, as well as changing through development. I will also extend these techniques into the investigation of activity during visual tasks. Despite being widely studied, visual networks such as those involved in face and object processing have had little formal network characterization. Given the importance of face and object processing in daily life, identifying properties of these systems has implications for understanding both typical and atypical functional brain organization in disorders like Autism Spectrum Disorder (ASD). In order to quantify these changes I introduce a novel approach to studying these brain dynamics I apply recent advances in spectral learning, an approach related to topics in applied graph theory and machine learning (Luxburg, 2007a) that allows the investigation of how networks change in time. Additionally I look at the spatial distribution of variability in these fluctuating systems, identifying a new differential network that seems independent of traditional functional networks, yet also predictive of function and development.

The specific aims of this thesis are:

1. To assess the organization of resting-state brain networks in healthy adults and children, as well as a small group of anesthetized primates, and determine network features that may be conserved across groups and species. (Chapters 4 and 5)

2. To build on the features observed in the prior aim to distinguish changes to brain organization during perceptual tasks relative to resting-state activity in both healthy children and adults as well as a cohort of children diagnosed with ASD. (Chapters 3, 4, and 5)

3. Develop a functional connectivity approach using spectral graph theory, simplifying and highlighting the most important network relationships present in the data. (Chapters 2 and 4)

During the completion of the studies in the present work the above aims became somewhat mixed as it was always useful to compare the activity at rest to what was observed during a perceptual task. Aim 1 is largely addressed in Chapter 4 where I not only investigate how spontaneous functional activity is organized, but introduce a novel method to identify how this brain organization changes dynamically in time. This advances our understanding of resting state activity as much of the current research in the field is looking at static networks derived from average activation patterns. It also provides a way to simulate possible disruptions to these networks, introducing an new method to probing translational links to other atypical groups. I continue to investigate resting activity in Chapter 5, as I compare the sources of variability in resting networks derived from both humans and primates. This work suggests a consistent and novel subnetwork in humans that does not appear to be present in this anesthetized primate group. After establishing these features of resting state networks, I investigate how they may reorganize in several groups viewing visual stimuli. In Chapter 4 I investigate dynamics in adults passively viewing faces and other objects, and observe a modest reorganization of dynamic organization, specifically the amplification of a ventral visual subnetwork that is matched to the presentation of images, and anti-correlated with a network related to resting state activity. In Chapter 5 this group is central to the development of differential networks, and I also look at children performing this task and identify a source of functional variability that appears to increase with age. I also compare these subjects to a cohort of drug using subjects viewing drug related parapher-

nalia. This group shows regions of the differential network with significantly higher degrees of variability. Much of my early work was looking at possible differences in children on the autism spectrum, and later on larger groups of subjects with autism from a public database. This largely yielded negative results, but I briefly go over these in Chapter 3, which also introduces some general concepts such as group averaged brain organization. Aim 3 was a methodological aim, and was central to the methods in Chapter 4.

1.2 Foundations of fMRI for neuroscience research

A general assumption of this work is that brain activity measured during fMRI through the BOLD signal is in fact an appropriate albeit incomplete and indirect representation of the underlying information as generated through neural activity. The investigation of how this neural activity is related to perception, cognition, and behavior is a central problem in neuroscience, and what follows is a brief review of the evidence of how fMRI has been functionally tied to other, typically more spatially localized, methods of measuring brain activity.

The development of fMRI

The development of neuroimaging techniques has progressed alongside the study of the brain itself. In the late 19th century, Angelo Mosso developed a technique known as 'human circulation balance', recording changes in blood flow in patients undergoing neurosurgery. He noted that there were differences in blood flow that occurred during cognitive activities. While this was the first recorded example of this technique, it remained unknown until recent discovery of his tools and manuscript.

In the early 20th century the neurosurgeon Walter Dandy developed ventriculography, which injected air into the lateral ventricles during anaesthesia, allowing x-ray imaging to better capture images of the ventricular system. Later Egas Moniz introduced cerebral angiography, which helped to visualize blood vessels in the brain through x-ray, though the positive contrast agents had negative effects. X-ray computed tomography (CT) scanning was developed in the 1960's and 70's by William Oldendorf, Godfrey Hounsfield and Allan Cormack. This was awarded the Nobel in 1979, and provided a much safer and easier alternative that continues to

be used, while still using radiation and often contrast agents.

Early blood flow maps were developed in the 1960's by Niels Lassen et al, using radioactive isotopes such as xenon, which produced images that could reflect brain activity from various cognitive tasks. This technique, combined with CT technology, led to the development of positron emission tomography by Edward Hoffman and Michael Phelps, first used in a human scanner in 1973.

Alongside these developments an alternative method was developed, magnetic resonance imaging(MRI). Rather than using some radioactive property, this used changes in signals made by hydrogen protons that are placed in a magnetic field. The 2003 Nobel was awarded to Paul Lauterbur and Peter Mansfield for the development of MRI. Over the years as the technology was developed it was noted that not only structural aspects of the body could be observed, but the kind of blood flows imaged through PET scanning were also measurable by MRI, leading to functional MRI more or less as it operates today. This has become the most widely used imaging technique in brain mapping due to its non-invasiveness and lack of any radiation. Briefly, neural activity leads to localized changes in blood flow, and within about two seconds oxygen-rich blood flows into the region displacing oxygen-depleted blood, gradually falling towards the typical baseline after about 6 to 12 seconds if activity has diminished. This diamagnetic oxygenated blood displaces the paramagnetic deoxygenated blood, which was introducing inhomogeneity to the local magnetic field, and this decrease leads to increases in image intensity which can be mapped onto the brain. The first *in vitro* demonstration of this signal change was done by Thulborn and Wright in the 1980's, and Ogawa et al showed that this was measurable in the rodent brain (Ogawa et al., 1992) using the dynamics of blood oxygenation as the contrast. The first human experiment was performed in 1991 by Kenneth Kwong using visual stimuli, published the

next year (Kwong et al., 1992). Kwong not only demonstrated the blood oxygen level dependent (BOLD) signal, but a direct blood flow signal that eventually led to methods known as arterial spin labeling.

The technology has continued to advance with stronger magnets and integration of associated technologies such as electroencephalography (EEG) and transcranial magnetic stimulation (TMS), but much of the progression in the field of fMRI has been through the development of algorithmic and computational techniques, leveraging modern computing power to apply increasingly sophisticated tools from statistics and machine learning. The present work is largely a continuation along those lines, investigating novel statistical applications to better understand whole-brain neural systems.

The functional basis for fMRI

The first fMRI studies identified large metabolic changes within occipital regions of the brain - regions that are well known to have visually evoked neural activity - and this was the first system to have a clear application of oxygenation of blood as a contrast in response to external stimulus. The visual system is a convenient region to test, as visually evoked neurons are among the most well studied due to relative ease of experimental design, easy access to the most accessible aspect of the CNS, the retina, and relatively easy access to primary visual cortex. This may also be reinforced by the general importance that vision plays in our everyday lives. This has led to a relatively robust computational understanding of sensory, particularly visual, systems.

Sensory and computational neuroscience took an enormous conceptual leap with the Nobel winning work of Hubel and Weisel (Hubel & Wiesel, 1959), which

used microelectrodes to identify the existence of a remarkable organization of function along gradients of cellular populations in the cortex. This organization appears to exist through laminar columns in a semi-discrete grid of population 'preferences' for different types of stimuli. Additionally there is a map from retinal input, through the thalamus, to V1 along the calcarine sulcus. This produces a radiation of inputs mapped and transformed through retina and thalamus away from this sulcus, with each hemisphere of the brain representing a contralateral hemi-field of visual space. This retonotopic map is fixed and evoked neural responses change as a function of input - ie as the eyes and gaze move around visual space.

fMRI encapsulates volumes of the brain that may contain tens or hundreds of thousands of individual neurons, depending on the spatial resolution of the imaging. If fMRI is capturing functional relationships between stimulus and neural activity, it should be possible to identify some of the fundamental functional organization already known to exist. Most importantly, much of this fundamental neuroscience work has been done in animals, and fMRI allows this knowledge to be applied to human subjects. Taking advantage of the known anatomical connections from retina, projecting into the lateral geniculate nucleus, then radiating into visual cortex, and using high field (3T) and localized high-resolution voxels, Schneider et al (Schneider et al., 2004) were able to image the laminar structure of the LGN. This anatomical body has 6 layered structures which receive distinct input from a hemifield of either the ipsi or contralateral eye, with dorsal layers having sustained color sensitive activity, with low contrast gain, and 2 ventral layers having more transient, high contrast and luminance sensitive activity. Thus different laminar profiles can and were observed through different stimuli presented to each visual hemifield independently. The relatively small superior colliculus was also partially mapped through a similar process (Sylvester et al., 2007), reproducing some aspects of

phase and movement often studied in this region in animals.

Those studies reproduced the most basic elements of vision, and confirmed in humans what was nearly universally accepted but not observed, that their visual system is organized as it is in other mammals. There is a long history of investigation of function of visual areas in animals, where function appears to increase in complexity from early V1 in the occipital lobe that is organized around more basic aspects of vision such as stimulus orientation, spatial frequency, motion, and color. As you move into V2 more complex visual relationships are encoded such as contours and patterns, V3 has more global feature and color mapping, as well as large scale motion. A widely accepted view of visual function argues that there are two distinct functional systems radiating from V1. This two-stream hypothesis (Goodale et al., 1991) suggest that a dorsal stream projects from the occipital lobe, through middle temporal, and towards the parietal lobe, while a ventral stream projects through V4 towards inferior temporal. These streams would largely serve different functions, where the dorsal system is concerned with processing visually guided behavior and motion, while the ventral system processes recognition or identification of visual stimuli. This leads to a conceptual delineation of vision into dorsal 'action, or where?' pathways and ventral 'identification, or what?' pathways. Recent reviews suggest that this is still a useful model, though the streams may not be as independent as once thought (R. D. McIntosh & Schenk, 2009). In any case, the ventral system is more relevant for much of the work in this thesis, where simple visual features are recognized and mapped in some feedforward manner, and upstream temporal areas encode very complex patterns, with evidence for various objects and faces having specific neural response patterns. Debate remains as to the specificity and sparseness of this encoding, although some studies suggest a remarkably sparse encoding (Quiroga et al., 2005), populations of neurons are still

likely required (Quiroga et al., 2008).

fMRI provides advantages in several respects to understanding the shifting anatomical and functional landscape encoding vision within the brain. First, these regions are largely discrete in function, but not necessarily through any obvious anatomical boundaries. Second, while animal studies are successful at investigations of fundamental properties of cellular function, which can be later confirmed in human subjects, more complex perceptual properties may only exist in humans and can only be studied through these non-invasive means. Dougherty et all (Dougherty et al., 2003) provided an early functional map of visual delineation between V1-3 using shifting stimuli, and showed a general consistency across individuals in how visual space is represented in the brain. fMRI has made it relatively easy to map additional visual regions (Wandell et al., 2007), such as the dorsal clusters V3a, b, V6, V7, and IPS, which tend to extend into parietal regions, often represent more peripheral regions (higher degree of tuning), are much more oriented towards motion, depth, spatial orientation, and eye-movement, are strongly modulated by attention, and damage tends to disrupt motion perception and visual/spatial attention. V5 in humans (MT or middle temporal in primates) tends to be completely motion sensitive, and can be difficult to track across individuals. Lateral clusters in occipital regions tend to have very large receptive fields and are broadly recognized to be involved in object and face perception. 'Early' lateral occipital cortex(LO1) continues to show selective responses to simpler visual features such as orientation selectivity, while LO2 begins to recognize more complex objects. More ventral clusters (V4, V8, VO1-2) become much more complex and difficult to study, though are involved in various ways of decoding color, complex shapes, and so on. These regions also tend to have less obvious homology with primate regions, where in early visual areas (V1-3 and V5/MT) there is reasonable

overlap. This could be due to some evolutionary differences, but may also be a limit of the tools used to identify these regions - a gap between functional fMRI and tools used to measure these areas in animals.

What about more discrete computational aspects, such as the functional columns that exist across cortex? These are typically on the order of hundreds of microns, at the edge of the higher resolutions available to BOLD signals. Ocular dominance columns were mapped in 2001 (Cheng et al., 2001), reproducing a well known feature where columns receive a balance of input from one, or both, eyes. A high strength (7t) scanner finally mapped the orientation columns found by Hubel and Weisel 50 years prior, and even imaged 'pin wheel' singularities known to exist between functional regions (Yacoub et al., 2008).

This evidence suggests that fMRI is able to identify many aspects of neural function at different scales, at least in sensory systems. The confidence in the underlying signal helps as studies step into more complex analysis that tends to abstract even further away from this signal, such as the network approaches used in this thesis.

The neural basis underlying the BOLD signal

While a functional basis underlying the BOLD signal may be well established, the specific nature of the neurobiology underlying this activity is still debated. Here I'll briefly outline some of the background relating BOLD to techniques that are more closely related to neural activity. While this issue is often secondary to the utility of BOLD in localizing function it is still critical to keep in mind the underlying basis for the eventual interpretation of many findings, particularly for research that may want to bridge gaps between findings from the fMRI literature and more localized

approaches to studying neural systems.

Considerable work has been done on correlating BOLD with simultaneously recorded electrical signals. This was largely done in animals and most notably from the lab of Nikos Logothetis (Logothetis, 2003; Logothetis et al., 2001). Briefly, electrodes placed in V1 were used to capture a wide range of frequencies of local neural activity while high resolution BOLD imaging of the area was also collected. This electrical activity can be decomposed through fourier analysis to identify individual spiking activity from other background activity composed of a superposition of many neurons activity. This lower frequency activity, often referred to as Local Field Potentials (LFP), can be influenced by neurons as far as 1 or 2 mm away, approximately the size of a voxel in BOLD imaging. After stimulation many neurons become active, but rapidly fall to a much lower firing rate. In contrast, the LFP tends to increase but gradually decrease to a constant rate of activity, before extinguishing at the removal of a stimulus. If a several second lag is taken into consideration this correlates well to the BOLD signal. Logothetis et al have succesfully created a simple model to predict BOLD signal given this LFP activity (Logothetis & Wandell, 2004). If spiking activity is understood as instantaneous activity resulting from a cellular integration of depolarization, LFP is a summation of this as well as all synaptic depolarization itself that does not necessarily lead to the spike itself.

Additional work has attempted to disambiguate spiking, LFP, and BOLD. One complex example used electrical records in cat visual cortex with simultaneously recorded local oxygen level measurement (Viswanathan & Freeman, 2007). Stimuli were used that are known to elicit synaptic activity at certain frequencies, but no spiking, and it was observed that LFPs were strongly coupled to changes in tissue oxygen in the absence of spikes, also supporting the closer coupling of LFP to BOLD.

While LFP is likely the closest analogue to the BOLD signal, clearly LFP is related to spiking activity, and a large enough sample of spikes would likely closely correlate to BOLD. A striking example of this was found by Mukamel et al (Mukamel et al., 2005), where two human subjects had spiking activity from multiple cells in auditory cortex recorded while listening to an auditory cue. This activity was convolved with a typical Hemodynamic response function that is frequently used to mimic the ideal dynamic of a BOLD response. These predicted BOLD responses were then highly correlated with BOLD from auditory cortex from an independent sample of subjects listening to the same cue.

A newer technology, 2 photon imaging, was recently used to attempt to address some similar issues. This used a complex setup that stimulated excitatory neurons in one region while recording BOLD in a distal targeted region (Lee et al., 2010). There was very clear BOLD activity in these target regions post stimulus, clearly showing that BOLD can be driven by neural excitation. They further claim this is evidence directly linking spikes to BOLD, rather than the summation of more distal synaptic activity. In response, Logothetis (Logothetis, 2010) writes a brief communication taking issue with this, stating that any local excitation could rapidly distribute into multiple neural circuits driving LFP activity in the target region, which was not denied by the authors. This study also attempted to observe the effect of stimulating inhibitory neurons, and found some evidence that this causes a local BOLD increase, with flanking voxels decreasing in magnitude.

These and other results all suggest that the BOLD signal is closely related to LFP activity, while also at least strongly correlated to spiking activity. The distinction lies in LFP reflecting 'inputs' and spikes reflecting 'outputs', which is ultimately important for a better understanding of additional analysis, including my own. BOLD activity may be correlated to these information carrying signals, but it is actually

driven by the changes in blood flow which is largely driven by astrocytes in a tripartite synaptic relationship (Figley & Stroman, 2011). Energy use, such as spiking activity, is not directly tied to blood flow, but rather indirectly related mainly through glutamate triggered calcium influx at the post-synaptic density, releasing a number of vasodilators. This typically increases blood flow over a relatively large area.

1.3 Characterizing the brain at rest

A growing body of research has focused on spontaneous changes in the BOLD signal during a resting state condition, where subjects are not engaged in any cognitive task and there are no experimenter controlled changes in sensory input. Changes from this baseline spontaneous activity during tasks are typically small (5% or less) compared to the total metabolic activity in the brain (Raichle et al., 2001). Importantly, this activity is not random and appears to reflect some underlying organization, which either re-organizes when engaged in a task, or specific task-related networks superimpose onto or modulate the underlying relations. As described by Fox and Raichle (Fox et al., 2005), evaluating and understanding this change from spontaneous to task related activity is an important unanswered question that likely requires novel approaches above and beyond voxelwise, seed, and independent component analysis based connectivity measures.

Spontaneous activity was first studied and formalized by (Raichle et al., 2001), and was based on earlier meta-analysis (Shulman et al., 1997), as a default or baseline state of activity. The regions showing consistent "deactivation" (i.e., activation during baseline compared to active tasks) included the posterior cingulate and precuneus, inferior parietal, and medial frontal regions. Further results often describe connectivity between regions through time series analysis, typically cor-

relations, reflecting some positive or negative colinearity between the respective signals. There is a lack of discussion within the literature as to the justification or weaknesses of using pearson correlations to derive brain organization. This may be a poor measure of information about some dependence between regions as it ignores the multiple influences imposed from areas across the brain, which is an implicit assumption in a network or systems approach.

Since first reported, the 'default mode network' (DMN) has been consistently observed, particularly using time series low-pass filtered below 0.1 Hz (Cordes et al., 2001; Fox & Raichle, 2007), though it also appears in block designs that are high-pass filtered. This network often has negative correlation with attentional networks (Broyd et al., 2009) suggesting a neural architecture mediating between periods of introspection and periods of attention and action. A number of other studies have gone on to perform similar analysis, typically either using a priori specification of seed regions (Fox, Corbetta, et al., 2006), or using some method of signal reduction across all voxels, such as independent component analysis (ICA) or principal components analysis (PCA) (De Luca et al., 2006; Kelly et al., 2008). These have found a number of candidate networks which appear to be synchronized during rest. These resting state networks have been observed across various conditions and samples, including anaesthetized monkeys (Vincent et al., 2007), have been placed under various classes, including (Fox & Raichle, 2007): DMN, visual, auditory, dorsal attention and ventral attention networks, and others. These resting networks include many regions that would typically have coincident activity during relevant perceptual or cognitive tasks, and there is evidence that there may be re-organization of resting state activity over development (Fair et al., 2007).

1.4 Development of functional organization, Autism Spectrum Disorder, and Face processing

Having established some resting 'baseline' network, the study will then contrast deviations from this baseline, in perceptual networks as well as in a cohort of subjects with Autism Spectrum condition (ASD). Studies in the present work will have subjects performing a passive face viewing task, where images of faces and other objects are presented with intermediate blocks containing no stimulus. This is a preliminary task made prior those that investigate how subjects perceive and compare the individual parts and overall configuration of a face, which were not analyzed in this work. The ability to correctly identify and process facial featural information is critical to social interaction, and disruptions to development of these processes can have considerable negative consequences to daily life, as in ASD.

While there is a considerable literature on various mechanisms of face and object processing, integrated network approaches to studying functional organization are relatively rare. Regions involved typically include the fusiform gyrus, superior temporal sulcus, and parts of the occipital cortex. These areas appear to form a functional core and may have different roles in various aspects of perception (Haxby et al., 2000), with fusiform representing identity and STS involved in changeable aspects of faces, though the regional integration and functional separation of these regions is still under debate (Cohen Kadosh et al., 2010; Collins et al., 2012; Sekuler et al., 2004). Other areas involved in numerous cognitive functions are also important for face perception and may interact with this core perceptual system, such as limbic regions for emotional context or parietal and frontal areas for processing eye gaze. This considerable imaging evidence in humans has supported single unit findings in macaque, where activity in superior and

inferior temporal cortex suggest that there are neurons that respond preferentially to face stimuli (Hasselmo et al., 1989; Perrett et al., 1982).

fMRI activation is typically greater for faces than other non-face objects in parts of this core system. Manipulations of face stimuli suggest that the neural systems in face recognition are much more sensitive to image inversion than object recognition (Yin, 1969). This inversion affects the 1st order information of the face; that is, the canonical order of face parts such as eyes above nose, nose above mouth. One explanation for the discrepancy is that relational processing is not as necessary for identifying objects, and provides a principal feature for separating faces from non faces, though not faces from one another as they have the same ordering of these features. 2nd order information, the spacing among face parts, is a candidate mechanism for face recognition and perceiving the changeable aspects of faces. This is important for recognition and non-verbal communication, and presumably less important for most object recognition. A debate as to the details of these aspects of perception and their independence in face and object detection continues. Network analysis of brain activity can help elucidate these largely behaviorally described manipulations.

Several studies have begun to look at developmental changes in functional connectivity in cognitive domains other than face processing. Fair et al. (Fair et al., 2009) looked at different age groups and found changes in regional clustering or modularity, with younger children having high correlations in close anatomical space, shifting towards functional integration across more distant regions in adults (more recently concerns with head movement in adolescents needs to be considered in interpreting these results; (Power et al., 2012)). Notably, Dosenbach (Dosenbach et al., 2010) used a multivariate statistical learning approach employing support vector regression to create a maturation index using resting state ac-

tivity from children to young adults. They showed that functional connections that changed with age tended to be more distal and along an anterior/posterior axis. There are few studies looking at typical developmental changes of visual perception (Cohen Kadosh & Johnson, 2007).

Autism is defined by deficits in impaired social interactions, impaired communication, restricted interests, and repetitive behavior (Rudie et al., 2012), and is increasing in prevalence. ASD is a disorder of neural development where alterations to synaptic coupling and organization lead to widespread functional changes. These changes are often described as relative increased local anatomical connectivity and decreased long-range connectivity, reflecting deviations from the typical developmental trajectory (Belmonte et al., 2004). An essential characteristic of ASD relevant to this study is an impairment in processing the emotional and social signals in faces. This may be from some dysfunction in aspects of face perception, through either overemphasis of individual face features, or reduced ability to process the overall configuration of a face.

Network analysis provides a means to characterize systems-level changes in development, evaluating connectivity both during rest, object, and face-viewing, and quantifying differences in typical and atypical populations. One hypothesis is that a support vector machine can predict condition and group from network statistics, providing a method of network quantification that incorporates subject-by-subject variability. Support vector regression will reproduce the maturation index from (Dosenbach et al., 2010), modeling ages in a large cohort of typical subjects ranging from children to adults, and then apply this to the smaller cohort diagnosed with ASD. This cohort is hypothesized to lie outside of the typical developmental profile, and changes in connectivity and network measures may reflect differences in brain development placing these individuals outside the modeled growth curve.

In a similar methodology, behavioral responses to visual feature discrimination will be modeled in both groups, and may reflect either a misallocation of resources, or the utilization of different sub-networks during discrimination.

1.5 Measures of similarity

This work takes a fairly analytical approach to understanding brain function. One goal is to not only describe the areas where function may be happening in a descriptive fashion, but to more specifically model how their dynamic activity is related in time and function. This approach is aided with some more advanced statistical and machine learning approaches, which help identify these dynamics in a rigorous manner. In this section I will briefly go over some of the basic methodological issues that need to be tackled prior to further analysis. A formal review of all the procedures taken in an fmri study would be unnecessary, and I won't go into detail on many of the more accepted techniques, such as those built into common packages like FSL.

Measures of correlation and dependence between time-series

Much of the analysis used in this research will depend on measures used to assess some kind of relationship between BOLD signal time-series recorded in two or more regions of the brain. Subsequent inference depends on this initial decision, so a review and understanding of some relative strengths and weaknesses will be useful before approaching more complex topics. In this section I will use some example data from several subjects in a continuous resting state scan, where no stimulus was presented for several minutes.

These metrics can be characterized by the nature of the relationship that they

measure - most notably whether the relationship is linear or non-linear, and additionally if the measure is conditioned on any additional data. The most classic, and far and away most applied in the fMRI literature, is the Pearson correlation coefficient, though many others have been developed and have had success in different applications.

Covariance, inner products, and the Pearson correlation

The covariance of two vectors measures the degree to which two variables change linearly, together. If one vector's values tend to increase as the others do, then the covariance will be positive. Otherwise, if one vector's greater values tend to correspond to the other vector's low values, then the covariance will be negative. As the degree of this relationship increases, the magnitude of the covariance increases, conversely as this decreases this magnitude will approach zero. If two vectors are independent, then they will have zero covariance. If two vectors have zero covariance, they may still be dependent (typically through some non-linear process). The covariance, or dispersion, matrix of vectors x and y is defined as: $\operatorname{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$. E[] is the expected value, or mean.

The units of measurement of this measure are those of X times Y. For the purposes of this discussion I'll be outlining population formula, though sample estimates of the respective variances are used in practice. In the context of linear algebra, covariance is closely related to an inner product in a euclidean space. This is often a useful interpretation for more geometrical approaches to information spaces, such as angles and norms of vectors, and allows for some important relationships often used in machine learning applications. Given vectors x and y, an inner product satisfies several rules:

Symmetry:

$\langle x, y\rangle = \overline{\langle y, x\rangle}$

Linearity:

$\langle ax, y\rangle = a\langle x, y\rangle.$

$\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle.$

Positive-Definiteness:

$\langle x, x\rangle \geq 0$ $\langle x, x\rangle = 0 \Rightarrow x = 0$

For random variables, $\langle X, Y\rangle := E(XY)$. Perhaps the simplest case is the inner product in a euclidean space, which is the simple dot product of two vectors. $x^T y = \sum_{i=1}^{n} x_i y_i = x_1 y_1 + \cdots + x_n y_n$. Inner product spaces have naturally defined norms, $\|x\| = \sqrt{\langle x, x\rangle}$. This is interpreted as the length of the vector, and allows the calculation of angles, distances, and so on. Many interpretations of machine learning techniques rely on finding geometric relationships in data which depend on these qualities. One important concept that I will not explore in detail is called the 'kernel trick', which is a widely used approach where a new function is used in place of some algorithm, such as covariance or inner product, generally for the purpose of finding non-linear relations. One important aspect of the selection of these kernels are that they satisfy rules of inner product spaces, triangle inequalities, and so on. When these are satisfied then we can effectively treat this information as if it were in our familiar euclidean space, and the location of values relative to each other has real information. Much of the work in the discovery of functional relationships in fMRI have used linear inner products, and it is worth keeping in mind that applying various kernel tricks may have some utility. In this thesis I will incidentally cover these briefly in Chapter 2, but applying kernels, for instance to the learning process in Chapter 3, remains a future direction.

An additional consequence of these geometric relationships is the triangle in-

equality:

$|||x, y|| \leq ||x|| + ||y||$

A simple distance metric is $\mathrm{d}(x, y) = ||x - y||$, or the euclidean distance. Another simple relationship within any inner product space is the calculation of the angle between two vectors. This would intuitively reflect some relationship, as a smaller angle would seem to suggest vectors pointing in a similar direction. Due to their geometric relationships, the calculation is straightforward: $\theta = arccos(\frac{x*y}{|x||y|})$. This leads to the most commonly used measure of dependence between timeseries, the Pearson correlation. This is the covariance, normalized by the product of each vector's standard deviation.

$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$

If this is compared to the calculation of the angle above, it can be seen that the right side of this equation is similar to the dot product of x and y over their norms. As stated earlier, the dot product of two vectors is equal to the product of their expected values. Moving the trigonometric function to the left side of the equation results in $cos(\theta)$, which maps the angle onto the range [-1,1], which is also the range of the Pearson correlation. $|x| = \sqrt{\langle x, x \rangle} = \sqrt{\mathrm{var}(X)} = \mathrm{SD}(X)$. When two vectors are orthogonal, the angle between them is 90°, and the correlation - or cosine of the angle - is 0. Many studies have used pearson correlations in contexts where a distance metric would be desired. Frequently this is done by transforming the Pearson correlation into a range that satisfies non-negativity, most often done with a distance correlation metric, which is simply $1 - \rho(X, Y)$. The best distance metric to use for various machine learning algorithms to capture information from fMRI activity is an open question, for the most part I will focus on various kinds of correlations between time series. This is principally due to the tremendous prior research that has largely focused on this measure, and the ease of application.

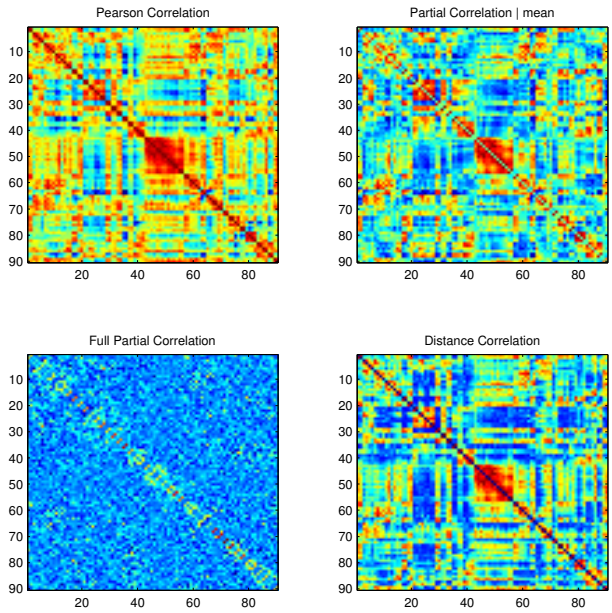


Figure 1.1: Examples of different correlation measures from one subject's time series. Each figure is a 90 by 90 matrix representing the individual relationships between all 90 brain regions studied throughout this work.

Partial Correlation

Partial correlations provide an interesting approach to measuring time-series dependence by controlling for some other random variables. This could be any variable that may be suspected to influence the results, and we would want to remove it due to our principle concern being the direct relationship between the two variables of interest.

The partial correlation ρ between X and Y, given one or more controlling variables Z, is the correlation between the residuals resulting from a correlation between X and Z, and Y and Z. One common application is to partial out a mean time series across all regions of interest in the brain. The use of this prior to analysis has caused some debate, as not conditioning on the mean tends to remove the anti-correlation between many regions, including influences from the default mode network. In this work, I typically calculate the partial correlation between

two variables, given all other variables. This is not the same as the global mean, as this seeks to provide as much information that is specific to two variables and not shared among regions of the brain synchronously. In practice, it is computationally efficient to calculate this full partial correlation using a matrix inversion of the covariance (or correlation) matrix. If the inverted matrix $\mathrm{P} = \mathrm{cov}^{-1}$ then the partial correlation between i and j is $\rho(X_i X_j) = -\frac{p_{ij}}{\sqrt{p_{ii} p_{jj}}}$. If there is some relation-

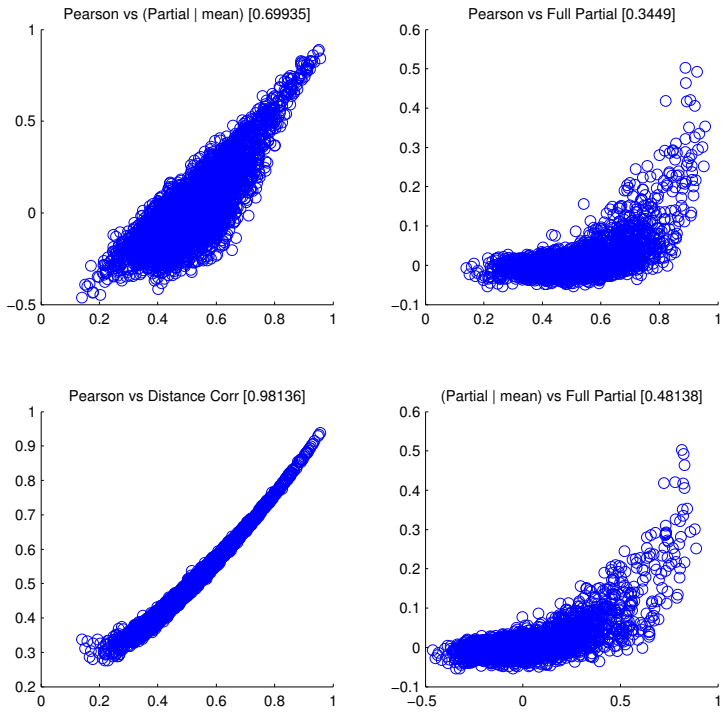


Figure 1.2: Scatterplot of different measures against one another, with r^2.

ship between three regions, where region A influences B, which influences C, full correlations would have high correlations in all pairwise connections. Partial correlation conditions each pair on other nodes - which can remove the 'extra' edge between A and C. This approach was described and applied in (Marrelec et al., 2006) where they suggest that this is a straightforward and data-driven technique that can be seen as more closely related to 'effective' connectivity approaches such as dynamical causal modeling, which are much more complex to calculate and re-

quire a number of prior assumptions. The method was further justified in (Smith et al., 2011) where it was shown through simulation work that partial correlation was more successful at recovering true underlying connectivity than Pearsons. In Figure 1.1 I show a comparison of several kinds of correlation computed from a single subject's timeseries across 90 brain regions. From this it can be seen that the matrix of partial correlations conditioned on the mean signal is qualitatively similar to the traditional Pearson's correlations. Conversely, there is a clear difference in the full partial case, where it appears many inter-regional links are removed. Figure 1.2 shows scatterplots between some of these kinds of correlation.

One significant problem often encountered by these researchers has been the difficulty in computing the inverse of the covariance matrix when the number of time series samples is smaller than the number of variables under consideration, for instance Marralec et al were only able to analyze a small subset of their data. In this work that will be a concern when looking at matrices derived from task-positive runs when subjects are viewing some stimulus category, as any particular stimulus will tend to only be viewed about 20 times, and the fixation runs used as intermediate resting states are viewed about 40 times. I typically use a partition of the brain into 90 regions, and calculating a covariance matrix from these data will result in unstable estimates that are singular - that is , they cannot be successfully inverted in order to calculate partial correlations.

I'll apply a simple shrinkage approach to improve this estimate, as developed in (Ledoit & Wolf, 2003) and applied in (Sch, 2005). This yields a well-conditioned positive definite covariance estimate that has guaranteed minimum mean squared error, and I apply a shrinkage factor, using the Ledoit-Wolf lemma, to calculate an optimal shrinkage intensity. $S^{\star} = \lambda T + (1 - \lambda)U$, where λ ranges from [0-1]. This serves as a convex combination of the sample covariance estimate U against the

target T, with a mixing parameter lambda to maximize the accuracy of the estimator. Essentially, as the sample gets larger, the estimate converges with the empiracle sample estimate. I use code written by and detailed in (Sch, 2005).

Distance Correlation

One additional metric that has recently gained attention is the distance correlation, sometimes referred to as a brownian distance. As mentioned, the classical measure Pearson's correlation measures the linear relationship between variables. Partial correlation measures a conditioned linear dependence. In 2007, Gabor Szekely (Székely et al., 2007) addressed this deficiency, where clearly (nonlinearly) dependent variables can have a 0 correlation. Given the random variables X and Y, first compute all euclidean distances in to the distance matrices A and B:

$a_{j,k} = \|X_j - X_k\|, \qquad j,k = 1,2,\ldots,n$

$b_{j,k} = \|Y_j - Y_k\|, \qquad j,k = 1,2,\ldots,n$ After centered these matrices by subtracting respective column and row means, and summing the matrix mean, calculate the distance correlation:

$\mathrm{dCov}(X,Y) := \frac{1}{n^2}\sum_{j,k=1}^{n} A_{j,k}\, B_{j,k}$

This effectively incorporates notion of linear dependence with dispersion of points, and in practice I used it as a quick measure of the non-linearity of timeseries relatedness. It is important to note a loss of sign or directionality in the relationship using this relationship (as distance is a strictly positive metric). In 1.1 and 1.2 it can be seen that distance correlation is qualitatively similar to the Pearson correlation, and in 1.2 it is nearly linearly related. If there were relationships that were not so linear, such as trivial functions like one region's timeseries being a square of another, then the distance correlation may reflect a relatively higher

value than appears in Pearson correlation. If there were complex relationships, it may be significantly higher in distance correlation while not being different from 0 in Pearson. This was only evaluated averaged across all subjects, and there may be subject-specific interactions where this distinction does appear. This metric remains a largely unexplored tool to investigate functional connectivity.

1.6 Graph theoretic approaches to connectivity

A graph theoretic approach to connectivity creates an abstraction of the brain with regions of interest as nodes, and the measure of connectivity relating pairs to each other as edges. A number of metrics have been used to characterize networks, such as node strength - a sum of all connections to any given node, or betweenness centrality - a reflection of how often a node lies on a shortest path between other nodes. There are also notable global measures, such as measures of small-world architecture, which can reflect a tendency for the network to organize around regions acting as hubs, which act as centers for other nodes to connect to and likely have high betweenness centrality themselves. Node modularity is another common measure, where the entire brain is partitioned into self-organized clusters or sub-networks. These various measures have been used successfully not only within neuroscience, but across numerous fields - in fact some of the most notable work has shown that there is a surprising consistency in some global network measures (Barabasi, 1999) across networks as disparate as metabolic networks and air travel. This 'non-random' aspect likely reflects important principles underlying information flow across real-world systems, where redundancy and relatively short travel times are at a premium, constraints that are also imperative to function in neural networks.

These graph heuristics have been developed alongside a rich mathematical framework, starting from Eulers use of graph theory to solve the Konigsberg Bridge problem with a formula relating the number of edges, vertices, and faces of a convex shape. A graph is best represented by a matrix containing edge values indexed between nodes, and this provides a great deal of statistical and linear algebra tools to employ in the study of graph properties. In particular is spectral graph theory, which studies summary properties of networks and their nodes through diagonalization of the system into its spectrum- the relationships among the eigenvalues and eigenvectors of that graph. As described by Chung (Chung, 1994), a major goal of graph theory is to describe principal structure in a graph through this spectrum, much as structural properties of a molecule can be inferred through its MR spectrum. I'll go into considerable detail to some topics related to modularity(or clustering) as well as spectral approaches to graphs in the following Chapter.

1.7 Concerns with head movement

Head movement has been a known problem for fMRI studies since early in its development. This can have a profound effect on the magnetic gradient being measured, as well as changing the position of the brain itself. Standard procedures have been designed for pre-processing of data prior to statistical analysis, these measure and realign the position of imaged volumes in space, and then frequently take some regression of the data against motion estimates to attempt to remove motion artifacts from the data. Recently (Power et al., 2012) it was shown that these common procedures were insufficient, particularly in younger subjects. Subjects with excessive average head motion are often discarded from analysis. Frequently, remaining subjects who otherwise have a low degree of mean movement

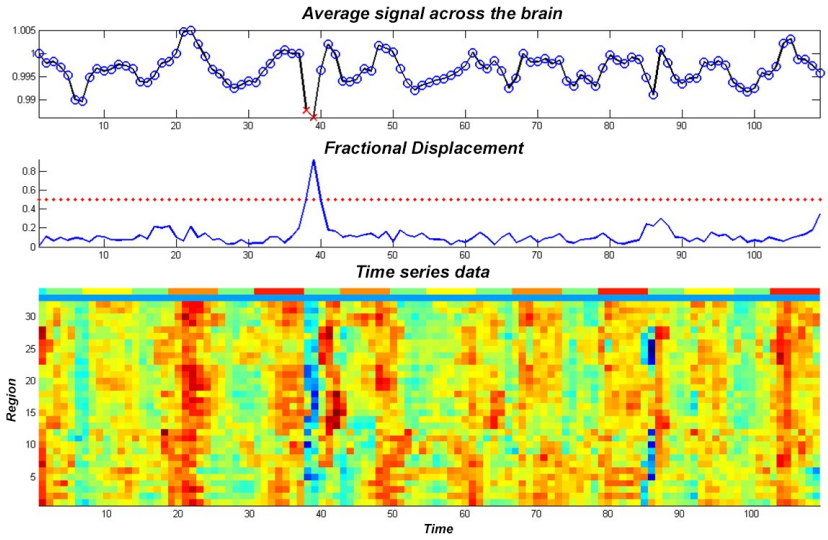


Figure 1.3: Here the mean time series(above), and the fractional displacement are plotted with a dotted red line at the cutoff point. Below is a region by region plot of the time series, with the top row indicating stimulus condition by color. There is a clear sudden movement at point 40, with an associated disruption to the time series. Red x's on the average signal are those frames that would be eliminated from analysis.

will have brief sudden movements, and recorded data at these points are completely in error. Powers et al suggest a procedure where a more thorough investigation of head movement is done on a frame-by-frame basis, and data from time points around sudden shifts in position are discarded from analysis. They show that these data from these times can have a large effect on functional connectivity, as even a few of these spurious movement artifacts can disrupt a real correlation, or possibly introduce a false one. In particular, they show that eliminating these movement points can alter the relationship between functional connectivity in children as compared to adults. Children tend to have more movement and this was likely contributing to some of the effect of studies seeing lack of functional connectivity. Removing these points increases the similarity between children and adults, and appears to recover some distal functional connectivity while still maintaining

a significant overall difference between groups. An application of the approach used by Powers et al was applied in most of the present work. In Figure 1.3 is an example from one subject who has a single rapid movement with an associated disruption in the time-series data. Removing these artifacts should improve estimates, and is especially important for developmental studies as in Chapter 3 or 5. Chapter 4 did not use this method to eliminate specific frames, as the analysis was specifically looking at a time-shifted correlation, and removing some number of frames would break the temporal assumption behind that approach. It should be noted that no children were analyzed in that work, so movement effects would be generally small, but it is something to keep in mind if in the future developmental questions are asked using this method.

Approaches to learning in biological and artificial systems

2.1 Introduction

Neuroscience has been applying methods from machine learning and statistical computing since these tools became available. There has also been considerable theoretical exchange between fields, as evidenced by concepts like artificial neural networks and conferences such as Neural Information Processing Systems. In this chapter I will cover a number of topics concerned with modeling unlabeled data, a field known as unsupervised learning. This is a contrast to several other kinds of artificial learning, one being supervised learning where some or all data have a corresponding label (such as having a pathology, something like age, or being a control) which are used to guide a search for distinguishing features, or reinforced learning where a process takes actions in some environment to maximize some kind of reward.

While the purpose of this chapter is mainly to discuss practical approaches to developing and applying clustering techniques, it is worth visiting how these artificial strategies may relate to cognitive and computational mechanisms in neural systems. Machine learning has been motivated by finding useful statistical approaches to understanding large amounts of data, and it may be informative to consider the evolution of brain systems in a similar context (Doya, 1999). Do the solutions that researchers find in understanding complexity and unknown structure have anything in common with the solutions the brain has found?

Reinforcement learning is probably the most clear example, and its development was guided by older observations from behavioral psychology. In the brain there are computational mechanisms underlying how reinforcement modulates dopamine (DA) in cortico-striatal systems. The first striking example comes from a series of studies by Schultz and colleagues (Schultz, 1998) on primate midbrain DA neurons. In conditioned motor tasks, these neurons initially respond to the reward on successful trials, but after learning the task these neurons shift to responding to the visual stimulus itself, rather than the reward. This ties directly into much of the computational literature, as temporal difference reinforcement models seek to devise a strategy that maximizes future reward (Sutton, 1998). The learning 'signal' in these models act exactly as DA neurons, at first linking motor behavior to the reward, and then tracking back in time to the causal mechanism underlying it. Another related example is in a task where stimuli are presented and subjects are required to make a binary decision to make a motor response, or withold that response, this is typically called the go/no-go task. Cortico-striatal inputs corresponding to motor and stimuli in go tasks are further excited by D1 receptors, whereas DA is inhibitory on nogo via D2. This differential will affect performance, and after a 'correct' response in the task dopamine bursts will reinforce the previous response. This results in learning during this phasic signal, strengthening the corresponding synapses, and facilitating selection of the previously rewarded response in future experiences with the stimulus. Likewise incorrect responses result in the opposite effect, strengthening nogo units which were being otherwise inhibited.

This would suggest that the basal ganglia are acting in a computational role to perform this reinforcement learning. Likewise, cortex has been suggested to act in a largely unsupervised manner, learning statistically optimal representations of the unknown state of the environment and its own behavior. This provides a mecha-

nism to learn behavioral 'habits' responding to events, and representations of the most important characteristics of stimuli, giving a shortcut to facilitate rapid action. This is evident in the ventral visual pathways discussed in the previous chapter, and these regions develop over time with additional experience and expertise with different stimuli. In the case of the go/nogo example, it has been suggested that sensorimotor pathways learn prior statistics of successful actions, at some rate slower than the basal ganglia guided learning (Frank & Claus, 2006).This can explain some results such as post-motor response activation in striatal units after action onset in well-learned tasks (Alexander & Crutcher, 1990), and patients with Parkinson's, a pathology of dopaminergic neurons, having relative ease with well-learned motor behavior. Dopamine also plays a role in 'activating' learned representations in cortex, as seen in addiction models as DA release in prefrontal cortex is necessary for drug seeking reinstatement (Kalivas & Volkow, 2005), but not its release in accumbens.

There are attempts to generalize unsupervised biological learning within a single framework, namely the free energy principle (K. Friston et al., 2006), which ties in closely to an even more general hypothesis of neural function known as the Bayesian brain. The free energy principal attempts to explain how biological systems minimize this energy among some number of internal states, where these internal states model belief about unknown or hidden states in their environment. This effectively minimizes self-information, or surprise, which is the divergence of the internal model from the external states. The objective is to maximize model evidence $p(s|m)$ or alternatively, minimize surprise $-logp(s|m)$ where the sensory state is evaluated given some generative model. This can be viewed as modeling the unconscious inference of perception, and an abstraction of neuronal processing. As the brain encounters the (unknown) world, it updates internal represen-

tations based on those sensations and how they compare to an internal representation. The generative models themselves that define the free energy function can be varied, but are typically hierarchical, like many systems in the brain. This can mimic message passing, and the movement of ascending prediction errors with their top-down predictions, a theoretical view consistent with the physiology of many sensory systems. As the most abstract levels at the top of the hierarchy make predictions about representations, lower, sensory, levels reciprocate with errors on those predictions. These unsupervised models would have the most in common with the underlying theory discussed in most of this chapter.

The cerebellum has long been understood to be involved in fine tuning of movement (Marr, 1969), though it is also innervated from large regions of cortex outside of motor areas. Fine motor control could be seen as an error-based learning task, which would correspond to supervised learning, as after initiation there is some target, and deviation from that can provide an error signal. A possible mechanism for this process was identified in the long-term depression of Purkinje synapses dependent on climbing fiber input which could provide some kind of error signal. Kitazawa et al showed this (Kitazawa et al., 1998) when they identify that the information content in spikes from these cells contain information about target direction at movement initiation, and error near end of the movement.

Besides these three general learning approaches, particular strategies to dealing with complex and high dimensional information in machine learning and statistics also seem to have biological 'implementations'. A typical example is the structure of the ear causing an incoming sound to produce a different frequency of oscillation along spatial regions of the membrane of the cochlea. This membrane has many hair cells, connected to nerve fibers, which are then each stimulated through a small range of frequencies of the incoming sound. This is effectively

a fourier transform from incoming mechanical wave over time into the frequency domain. There has also been considerable interest in understanding the underlying dimensionality and functional organization of neural populations and the information contained in their activity. The clearest example is the encoding of the geometry of an animals spatial environment in hippocampal place cells(and the related grid cells in nearby enthorhinal cortex), where a proportion of neurons becomes active in a specific field of space in the environment (O'Keefe & Dostrovsky, 1971). The population of cells, which is flexible to new environments, is then able to represent information about the entire area to produce a cognitive map of the environment. Given the role the hippocampus plays in memory it is suspected that this map is important for creating spatial context for memory formation. Place cells are able to maintain activity after changes to the environment (Moser et al., 2008), which may suggest a role in the ability to complete memory patterns only using fractions of information. There are also hypothesized mechanisms using this system that act to separate patterns, where the dentate gyrus relays sensory information to place cells, and the constructed maps ultimately relate back to higher order patterns in cortex (Rolls, 2013). I'd suggest this technique is closely related to statistical practices such as principal components analysis, where a simpler low dimensional representation of information is stored in a high dimensional, neuronal, space. This embedding in lower dimensions is itself informative, and some might say 'emergent', and as we'll see can guide techniques such as clustering. This general computational approach to lower dimensional representations for encoding in cortical areas will likely continue to gain attention (Barbieri et al., 2014; Citti & Sarti, 2014).

The brain itself appears to have multiple approaches to modeling information, and it is no surprise that there are numerous ways for investigators to, in turn,

model the brain. The motivation for the remainder of this chapter is to introduce and formalize some of the established techniques and examine how they are related. These typically model hidden or latent structure in the data, either through identifying cluster centers (kmeans), additionally adding mixture weights and modeling variance about those centers(Gaussian Processes), or finally modelling the dynamic nature of changes in latent states and influence of observations over time (Hidden Markov Models). I contrast those traditional approaches with another class of algorithms from multiple disciplines that looks for useful embeddings of information in lower dimensions, and identifies structure in this simplified space to then identify cluster centers, their variance, and ultimately their dynamics. This new approach is computationally more efficient, and results in a tractable system to learning latent parameters and a successful strategy that unwinds the complex picture given to us from whole brain analysis in fMRI, and I apply that to find the results in the next chapter.

2.2 Clustering

There are countless examples of unsupervised learning in the neuroscience literature. In previous lab work prior to dissertation research I spent a considerable amount of time identifying and characterize spikes from populations of neurons recorded from extracellular electrodes. The system capturing these signals had built in software that would take spike shapes, project them into lower dimensions using principal components analysis, and then run a clustering technique - k-means - to try to identify some unknown, or latent, structure separating different neurons. This process could be largely automated and literature often did not go into detail on how spikes were discovered - even though small changes to some

parameters could make very considerable differences in results.

With my current research I've encountered a similar problem, though on a very different scale. fMRI timeseries spatially encompass the entire brain and contain large amounts of information, but how best to manage and characterize this information is still being established. Significant work has gone into understanding how different clusters, or sub-networks, seem to preferentially communicate and organize their activity. Clustering is a very difficult problem, and the notion of a cluster or partition itself is poorly defined (Estivill-Castro, 2002). Being unsupervised, it is a problem where some intrinsic aspect of the data is being modeled, with no outside information or labels to help guide the process. A full review would be impossible, but I will go through a number of the most common approaches which will be important to the following work. These all typically try to organize observations, such as regions of the brain, according to some measure of how related they are. For instance, bilateral regions from left and right lobes typically have similar time-courses, which would be strongly correlated and subsequently likely to be identified in the same cluster.

Hierarchical clustering is one of the simpler and older ways to assess data organization. The process does not create a single cluster assignment, but a multilevel tree. As one moves through the tree, distinct clusters at one level are joined at the next. Tree construction requires a similarity (or dissimilarity) score between every pair of variables, which in my case is typically regions of the brain. The typical correlation or partial correlation matrices will work, and these are used to create linkages between sets of variables to determine the order that sets are combined. There are a number of criteria that can be used, such as the minimum or maximum among all distances between sets, or a factor tied to the decrease in variance when merging clusters. This results in a tree structure that can be 'pruned' at any level

to create a desired number of clusters. I frequently use this approach as a reliable and minimally complex way to assess brain organization, as the only decision beyond cluster number is the choice of distance and linkage function, and results are deterministic. This by no means suggests it is correct, and is often weak to outliers which form clusters themselves, causing other clusters to merge.

One popular approach in the fMRI literature has been evaluating brain organization through the lens of graph theory, where regions are nodes in a graph that are connected by edges (Bullmore & Sporns, 2009). These edge values are either binary or determined by a 'weighted' similarity metric such as a correlation structure. Many algorithms were developed in the context of binary connections between nodes, and particularly when moving to weighted graphs this approach is often more of an abstraction of statistical and machine learning methods. To produce a graph, correlations are often thresholded at some value to produce more graph-like structures. Many important results from this literature come from the surprising tendency for complex systems to self organize into modules, or clusters, around the most important nodes, referred to as hubs. More generally, a network with community structure can be grouped into modules containing more densely connected sets of nodes. Numerous algorithms can be used to determine this structure, such as hierarchical clustering.

2.3 K-means

It will be useful to start formalizing some of the variables and methods used in cluster analysis, beginning with K-means. Given a dataset consisting of N brain regions $x_1, ..., x_t$, our random variable, in D dimensions, and it is desired for these regions to be split into K cluster assignments. As discussed, typically a measure

is used for clusters that minimizes inter-cluster distances compared to outside distances. If cluster centers are represented by a D-dimensional vector μ_k for each cluster, then a minimization of the sum of squares from data point distances to a closest vector μ_k is made. Using a binary indicator r_{tk} associating a data point to a cluster, I can define the objective function

$$J = \sum_{t=1}^{N} \sum_{k=1}^{K} r_{tk} ||x_t - \mu_k||^2 \tag{2.1}$$

.

The goal is to minimize J through finding values of r_{tk} and μ_k. This is done by initializing any value for μ_k, and keeping this fixed, minimize J with respect to r_{tk}. Next, minimize J with respect to μ_k, keeping r_{tk} fixed. Then repeat this two step process, at one point assigning clusters, and the next locating cluster centers, until some number of steps or convergence of the objective. Each iteration will necessarily lower or converge on a value, but the value reached may be a local minimum. This means that there may be many combinations of clusters that are solutions, but each application of the k-means algorithm could reach a different solution. Frequently many passes of the algorithm might be performed, looking for a most minimal objective. J is a linear function of r_{tk} and variables are independent, so the n^{th} data point is simply assigned to the closest cluster center. Updating μ_k is quadratic, and is minimized throughits derivative

$$2 \sum_{t=1}^{N} r_{tk}(x_t - \mu_k) = 0 \tag{2.2}$$

$$\mu_z = \frac{\sum\limits_{t} r_{tk} x_t}{\sum\limits_{t} r_{tk}} \tag{2.3}$$

The interpretation is simply that μ_k is set equal to the mean of all data within that cluster, the numerator effectively summing the data, and the denominator counting the number of variables. The steps of the k-means clustering correspond closely to something called the EM algorithm, where updating r_{tk} can be considered the expectation step, and updating μ_k the maximisation step. The EM algorithm is a useful method for finding maximum likelihood solutions in problems with unseen, or latent, variables.

2.4 Gaussian Mixtures

An extension of this approach is done by modeling more than just cluster means. A useful approach is to look at gaussian mixture models which also incorporates information about the variability about these means. This can be viewed as a superposition, or linear combination, of individual gaussians. Each of these could be considered the distribution underlying the respective class, and by adding more gaussian distributions, effectively any continuous distribution could be modeled. Given K Gaussian distributions $\mathcal{N}(x|\mu_k, \Sigma_k)$ the mixture is

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k) \tag{2.4}$$

where π_k is a weight or mixing coefficient which should sum to 1. This can be viewed in terms of discrete latent variables, where I'll introduce an additional K-dimensional binary random variable z, which has one dimension = 1 and all others = 0. The marginal over z is set equal to the mixing coefficient. The marginal of x is obtained by summing the joint over all z, which simply results in a Gaussian mixture as in (2.4). This creates a new representation of the marginal that allows

for a latent variable z_t for each data point, and allows some applications through use of the joint $p(x, z)$ with the EM algorithm. This leads to the introduction of an objective function to maximize, similar to (2.1). From (2.4), the log likelihood is given by

$$\ln p(X|\pi, \mu, \Sigma) = \sum_{t=1}^{N} \ln \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k) \tag{2.5}$$

. Taking the derivative with respect to μ_k to 0,

$$0 = -\sum_{t=1}^{N} \frac{\pi_k \mathcal{N}(x|\mu_k, \Sigma_k)}{\sum_j \pi_j \mathcal{N}(x|\mu_j, \Sigma_j)} \Sigma_k (x_t - \mu_k) \tag{2.6}$$

The fraction summed over N happens to be the conditional $p(z_k = 1|x)$ and if π_k is a prior over z_k, this is a posterior after seeing x. If this posterior is referred to as $\gamma(z_{tk})$, and defined through $N_k = \sum_{t=1}^{N} \gamma(z_{tk})$, after multiplying by the inverse covariance Σ_k^{-1}

$$\mu_k = \frac{1}{N_k} \sum_{t=1}^{N} \gamma(z_{tk}) x_t \tag{2.7}$$

. That is, the mean for the k^{th} Gaussian is the weighted mean of all points, with the weight determined by the posterior that this component was responsible.

A similar process, finding the derivative with respect to Σ_k yields

$$\Sigma_k = \frac{1}{N_k} \sum_{t=1}^{N} \gamma(z_{tk})(x_t - \mu_k)(x_t - \mu_k)^T \tag{2.8}$$

Next, I maximize with respect to the mixing coefficients, which are required to sum to one, which subjects the function to a langrange multiplier (details omitted here), ultimately yielding the fraction for mixing components

$$\pi_k = \frac{N_k}{N} \tag{2.9}$$

The k^{th} mixing component is the average over the posterior of that component. These steps produce a way to implement EM to update the Gaussian mixture estimate, where the Expectation step is done by estimating the posterior $\gamma(z_{tk})$ from current parameters, and then these posteriors are uses to update the parameters in (2.7,2.8,2.9). Again, similar to Kmeans, this process is guaranteed to increase the log likelihood, and the process is typically stopped as the change in log likelihood is below some value.

2.5 Hidden Markov Models

While the techniques discussed so far can provide considerable insight into brain organization, our data is sampled in time and it may be false to assume that it is identically distributed sequentially. This is relatively common in various measurements, from time series in general, or certain kind of sequentially related information such as text or DNA, and I'll be applying similar methods to investigating changes in functional brain organization over time. The joint distribution of a series of data over time T from a Markov model is

$$p(x_1, ..., x_t) = \prod_{t=1}^{N} p(x_t|x_1, ..., x_{t-1}) \tag{2.10}$$

If observations only depend on the immediately preceding observation

$$p(x_t|x_1, ..., x_{t-1}) = p(x_t|x_{t-1}) \tag{2.11}$$

It is possible to consider higher order markov chains, which model earlier parts of the time series, but doing this grows model parameters exponentially. A more useful way to add flexibility to the model is to add discrete latent variables, as was

done for Gaussian mixture models, where each observation has a corresponding latent variable z_t. A Hidden Markov model is an extension of mixture models where the mixture component at each observation is not independently selected, but depends on the previous observation. If these latent variables are K-dimensional binary indicators of state, they correspond to a table of conditional values $p(z_t|z_{t-1})$ T, known as a transition probability matrix. The conditional distribution is then

$$p(z_t|z_{t-1}, A) = \prod_{k=1}^{K}\prod_{j=1}^{K} A_{jk}^{(Z_{t-1,j})(Z_{tk})} \tag{2.12}$$

Where A is a table of numbers who's elements are the transition probabilities for moving between states. This table has elements $A_{jk} = p(z_{tk} = 1|zn - 1, j = 1)$. The starting state has no parent, so its marginal is represented by a vector of probabilities π. Additionally, the observation or emission probabilities are defined as

$$p(x_t|z_t, \phi) = \prod_{k=1}^{K} p(x_t|\phi_k)^{z_{tk}} \tag{2.13}$$

which could be from any distribution governed by parameters ϕ. The full joint distribution over latent and observed variables is

$$p(X, Z|\theta) = p(z_1|\pi)[\prod_{t=2}^{N} p(z_t|z_{t-1}, A)] \prod_{m=1}^{N} p(x_m|z_m, \phi) \tag{2.14}$$

$\theta = \pi, A, \phi$.

After observing some data, I can again use EM for maximizing likelihoods, though now a different approach is required as complexity will otherwise grow exponentially with the length of the time series. It is possible to take advantage of the conditional independence of the states to reduce the cost of inference. After

initializing θ^1 I begin an expectation step.

$$\gamma(z_t) = p(z_t|X, \theta^1) \tag{2.15}$$

$$\epsilon(z_{t-1}, z_t) = p(z_{t-1}, z_t|X, \theta^1) \tag{2.16}$$

Where γ and ϵ are the posterior and joint posterior of latent and successive latent variables. Each marginal state γ can be a vector of K positive values summing to one, and ϵ is a matrix summing to 1. An efficient and frequently used procedure for evaluating these is known as the forward-backward algorithm (Rabiner, 1989) or Baum-Welch algorithm (Baum, 1972). The evaluation of latent variables is independent of the form of the emission. We are interested in finding the posterior $p(z_t|x_1, ..., x_t)$ given the observed data.

$$\gamma(z_t) = p(z_t|X) = \frac{p(X|z_t)p(z_t)}{p(X)} = \frac{p(x_1, ..., x_t, z_t)p(x_{t+1}, ..., x_t|z_t)}{p(X)} = \frac{\alpha(z_t)\beta(z_t)}{p(X)} \tag{2.17}$$

Where α represents the joint of observing all data and the current state, and β represents the conditional of future data given the state, and were separated using their conditional independence given the current state. These are also vectors of size K. $p(X)$ is conditioned on θ and represents the likelihood. It is possible to define recursive relations of these in terms of previous states

$$\alpha(z_t) = p(x_t|z_t) \prod_{z_1^{n-1}} \alpha(z_{t-1})p(z_t|z_{t-1}) \tag{2.18}$$

The first step of the recursion is initialized

$$\alpha(z_1) = p(z_1)p(x_1|z_1) = \prod_{k=1}^{K} \pi_k p(x_1|\theta_k)^{z_{1k}} \tag{2.19}$$

Each step ends up involving multiplication of a KxK transition matrix in $p(z_{t+1}|z_t)$, so the overall cost is $O(K^2N)$. A similar process can find a recursion relation for β

$$\beta(z_t) = p(x_{t+1}, ..., x_t|z_t) = \sum_{z_{t+1}N} \beta(z_{t+1})p(x_{t+1}|z_{t+1})p(z_{t+1}|z_t) \tag{2.20}$$

Each step uses the observation x_{t+1} through the emission $p(x_{t+1}|zn+1)$ multiplied by the transition $p(z_{t+1}|z_t)$ The initial condition in this case is the value at N, which from (2.17) could simply be viewed as the vector of ones as the limiting case as α covers the entire joint distribution. Next I'll evaluate the matrix in ϵ which can be put in terms using α and β

$$\epsilon(z_{t-1}, z_t) = \frac{\alpha(z_{t-1})p(x_t|z|n)p(z_t|z_{t-1}\beta(z_t)}{p(X)} \tag{2.21}$$

This produces a new set of parameters which can be used to evaluate a maximization step using these terms. This produces updates

$$\pi_k = \frac{\gamma(z_{1k})}{\sum_{j=1}^{K} \gamma(z_{1j})} \tag{2.22}$$

$$T_{jk} = \frac{\sum_{t=2}^{N} \epsilon(z_{t-1,j}, z_{tk})}{\sum_{k=1}^{K}\sum_{t=2}^{N} \epsilon(z_{t-1,j}, z_{tl})} \tag{2.23}$$

To maximize ϕ an update similar to that of a standard mixture for independent data can be applied, where γ acts as the posterior as in (2.7,2.8).

It is typically desirable to have an estimate for the most probable sequence of states given observations, and the parameters from an HMM, which is not the same as identifying the most probable state at each point. I'll just briefly mention

the most common approach to evaluating this, the Viterbi algorithm, which saves computational cost by using a dynamic programming approach to searching all the possible paths. For a more complete reference for all these steps refer to (Bishop, 2007).

2.6 Dimensionality reduction and Spectral Theory

Next I will introduce an important class of analysis that applies to theories involving mathematical operators in linear algebra and linear equations. This is an important branch of applied mathematics with a long history in physics and statistics. Some insight into motivations behind its use in physical systems might be gained from a well known article "Can one hear the shape of a drum?" in the American Mathematical Monthly. The question was, could the shape of a (theoretical) drum be uniquely predicted if the frequencies were known? The frequencies happen to be the eigenvalues of the laplacian of the membrane, where the laplace operator is a differential in space. It was eventually shown that some different shapes did have identical eigenvalues, known as their spectra, though these drums do have the same area and perimeter. In fact the area and perimeter could be approximated through analysis of the distribution of eigenvalues. So in this sense the spectra seem to map information from one space to another, similar in some respects to a fourier transform.

More generally, spectral algorithms study information contained in eigenvalues(or singular values) and eigenvectors of various kinds of matrices, or graphs. Outside of theoretical physics, there were important applications using these approaches that have become commonplace - first and probably most notably principal components analysis(PCA). PCA is a relatively simple approach to dimension

reduction which has been applied for nearly a century. PCA finds an orthogonal basis for some dataset $X \in \mathbb{R}^{n*t}$ that seeks to maximize the variance projected onto that basis. Despite being a complex, nonconvex problem, PCA has an exact solution (Eckart & Young, 1936). Importantly, it can be calculated with standardized linear algebra methods, the top k vectors in a dataset are given by the top k eigenvectors of the covariance matrix, and the corresponding eigenvalues are proportional to the variance explained by that vector. Singular value decomposition(SVD) is a closely related method than can also compute these principal vectors, without first computing the covariance. This can be put in terms of a low-rank matrix approximation, that is, the decomposion of data matrix X such that $U\Sigma V^T = X$, and then keep the top k singular vectors $X_k = U_k\Sigma_k V_k^T$, which reconstructs the data in a least squares sense at a given rank.

These methods have been commonplace in statistical applications for decades. The dimensional reduction and low rank approximations discussed above can also improve clustering methods, as they remove noise and reduce the effect of the curse of dimensionality, where the volume of a high dimension space requires exponentially more sampling to cover it. There was a resurgence of activity in the machine learning community to applications of spectral theory to clustering and graph partitioning. An example was in (Ng, Jordan, 2002), where rather than using a simple covariance or correlation matrix, a more general affinity matrix from the data is constructed using a gaussian kernel

$$A_{ij} = exp(-||s_i - s_j||^2/w\sigma^2) \tag{2.24}$$

. With A, a laplacian is related to that used in the physical membrane problem,

$$L = D^{-1/2} A D^{-1/2} \tag{2.25}$$

Where D is a diagonal matrix with the sum of rows of A along the diagonal. Now, after performing an eigendecomposition as in PCA to form a low dimensional projection, they also normalize to have row unit length (projection on a sphere). Clustering is done in this space, via K-means or any other algorithm. These steps tend to have much more success in clustering distributions, particularly those which have complex non-linear qualities rather than spherical or gaussian characteristics.

A closely related perspective to spectral clustering can be gained by viewing the affinity matrix as a graph, and considering a process of random walks on this graph. A cluster could be seen as a partition where the random walker tends to stay for long periods before moving to another section of the graph. Formally, the transition probability is proportional to the edge weights, and given by $P = D^{-1}A$. This random walk will have a unique stationary distribution $\pi = (\pi_1, ..., \pi_n)$ where $\pi_i = d_i/vol(A)$ and vol(a) is the total weight in the graph. There is a close relationship between this transition graph and the laplacian, as L = 1 - P. There is a corresponding relationship between respective eigenvalues, with the largest in P matching the smallest in L (Luxburg, 2007a).

Along these lines some additional, and perhaps more fundamental, relationships in dimensionality reduction have been established. Coifman et al (Coifman et al., 2005) discuss a framework that generalizes these results to show that eigenfunctions of Markov matrices can bridge the local transition probabilities into a macroscopic description of the system, mimicking physical integration of small

movements into global changes. They show that a diffusion distance is an important geometric quantity that can embed data into a euclidean space according to a diffusion metric. They establish a family of diffusion maps

$$k_\epsilon^\alpha(x,y) = \frac{k_\epsilon(x,y)}{p_\epsilon^\alpha(x)p_\epsilon^\alpha(y)} \tag{2.26}$$

Where the numerator is effectively the weight or affinity function, $p(x) = \frac{e^{-U(x)}}{Z}$ which is a steady-state Boltzmann density. When α = 0, this is the classical normalized graph laplacian, when α = 1, this is the Laplace-Beltrami operator(heat kernel smoothing), and when α = 1/2, this is related to the backward Fokker-Planck operator, which is a partial differential equation describing time evolution of probability distributions under random forces, such as brownian motion. They suggest that diffusion maps of points sampled from dynamic systems may be a tractable method to recovering their lower dimensional eigenfunctions, rather than using numerical solutions of differential equations.

2.7 Spectral Learning

While the earlier methods to latent variable discovery, kmeans, gaussian mixtures, or Hidden Markov models, have all been incredibly useful and successful at modeling various systems, the underlying learning problem is not completely solvable - local minima may always exist, and results can be sensitive to selection of initial states. Recent approaches have tried to make separation assumptions about the underlying distributions. For instance Vempala and Wang (Vempala & Wang, 2004) show that it was possible to use a spectral approach to rapidly learn a mixture of k spherical(uniform variance) Gaussians. The main step is to project to a subspace

spanned by the top k singular vectors of the sample matrix. They show that after projection, the separation between mean vectors of the underlying distribution is preserved while the radius of the distribution drops by a factor proportional to k. This effectively amplifies the ratio of separation while lowered the dimension. The results also hold with any moderately isotropic distribution. Work along these lines has largely concentrated on what the magnitude of separation between distributions needs to be for efficient learning algorithms, for instance bounds can be set on the magnitude of variance (or radius) for a correctly classified Gaussian.

There are clearly close relations between this approach to learning a gaussian mixture model, and results from topics such as spectral graph theory as discussed earlier. Besides the general utility in clustering in lower dimensional spaces, the spectral gap of a markov chain can be defined as the difference between its two largest eigenvalues and puts a bound on the mixing time for that chain. IE the 'bottlenecks' present in a graph could be seen as larger distances between clusters, and increase transit or mixing time. The clear relationship in discovering latent distributions when comparing kmeans, mixture models, and HMMs above, would suggest that there may be a method to use these various insights in spectral embedding to not only improve classification performance, but learn dynamics as in HMMs.

Hsu et al made an important advance in this direction (Hsu et al., 2012). They combined these insights in separation of mixtures with results from several closely related fields. One, subspace identification, used spectral approaches to discover latent relationships in linear dynamical systems, which are related to HMM's but assume Gaussian distributed latent states as well as additive noise independent of these states. They combine this with insight from the use of Observable Operator Models (Jaeger, 2000), or the closely related Predictive State Representations

(Littman & Sutton, 2002), which developed non-iterative asymptotic approaches to HMM inference which represent sequence probability as the product of matrix operators. These techniques main learning procedure is done by using spectral decompositions of correlations between past and future observations. This is a small but important shift from most work that uses 'stationary' correlations between time series, which makes up the majority of work on functional brain organization.

This effectively yields a simple and rapid method to learn HMMs, in a strikingly different way from EM as outlined above, using some separation conditions similar to those studied in gaussian processes. This can approximate the joint distribution over observation sequences (2.28, 2.29), as well as the conditional distribution of future observations on some history (2.39), and the error is asymptotically bounded. This results in a model of dynamics that does not explicitly model transitions between states, but does linearly relate to HMM's.

Basic Algorithm: Learn-RR-HMM(k,N)

from (Siddiqi et al., 2009)

1. Compute empirical estimates $\hat{\vec{P}}_1$, $\hat{P}_{2,1}$, $\hat{P}_{3,x,1}$ $\forall x$ directly from the data.

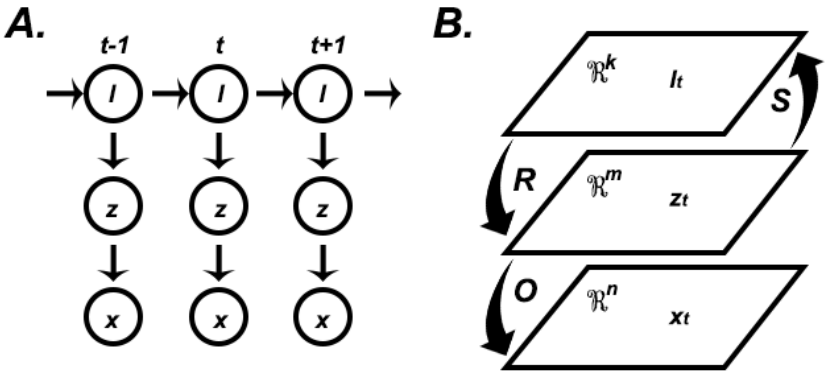
2. Use SVD on $\hat{P}_{2,1}$ to compute $\hat{U}$, the matrix of left singular vectors for each of k largest singular values.

3. Compute the model parameter estimates $\hat{b}_1$, $\hat{b}_\infty$, $\hat{B}_x$

This algorithm is idealized for sampling N triplet sequences from a dynamical system. In practice this is more often 1 or more long time-series which can also be used to develop estimates, and this extension is also detailed by Siddiqi et al. For practical consideration of these estimates see below in implementation details, and in the appendix for some code examples.

This will continue from the definitions of a transition matrix describing m states $T \in \mathbb{R}^{mxm}$ (2.12), and observation matrix $O \in \mathbb{R}^{n*m}$ (2.13), as well as an initial state

Figure 2.1: A. Graphical representation of the reduced rank HMM in time as outlined by Siddiqi et al. Observations x are determined through the latent state z, while l is a k rank projection of that hidden state. B. Parameters in the model and the variables they act on. O is the transition probability of observation space given the state z, while R is propagating the low rank state to the hidden state, while S is projecting the hidden state to the low rank state.



$\pi_i = p(z_1 = i)$. The alternative view of computing sequence probability from observation operators, for observed variables in vectors x_t, for example our n regions of the brain

$$A_x = Tdiag(O_{x,1}, ..., O_{x,m}) \tag{2.27}$$

where for any time t the joint probability is calculated:

$$p(\vec{x}_1, ..., \vec{x}_t) = 1_m^T A_{x_t} ... A_{x_1} \pi \tag{2.28}$$

Where diag(O) is a diagonal matrix with the respective observation probabilities, and 1_m^T is a diagonal ones matrix. Hsu et al assume that the probability of initial states $\pi > 0$ and that the rank of the transition matrix is m, that is there are not some states which are somehow mixtures of other states. This assumption was relaxed in (Siddiqi et al., 2009) where the underlying system as modeled in T may have low rank, k, where $T = RS$, $R \in \mathbb{R}^{m*k}$ and $S \in \mathbb{R}^{k*m}$. This casts the

initial state distribution into a lower dimensional state, $\pi = R\pi_l$, and they refer to this reformulation as a reduced-rank HMM. R can be thought of as acting on the discrete latent space, while S acts on the low-rank continuous latent space. S projects z_t to l_t, while R propagates l_t to z_{t+1} (See Figure 2.1 for a graphical model). This provides an alternative distribution to (2.28), and they show that a reduced-rank HMM with rank k is actually more expressive than a k-state HMM, and it is able to more smoothly model transitions between states. Rank-k RR-HMMs (which can have m $>>$ k states) can produce sets of predictive distributions not possible in k-state HMMs. This is an important quality for systems that have smooth intermediate regions between states, an example used by the Gordon lab in presentations is learning the motion of a pendulum(through video recording), a standard HMM can learn a set of velocities and directions, but a reproduction from the generative model yields a very 'jerky' system jumping between these states. This reduced-rank technique is able to represent combinations of states, which through simulation can yield smooth trajectories from the same number of basis states, or the same complexity model. This is almost certainly an important feature in the brain, which is likely to have transitions between states rather than shifting between sets of discrete modes.

$$p(x_1, ..., x_t) = 1_m^T R W_{x_t} ... W_{x_1} \pi_l \tag{2.29}$$

This formulation avoids using EM and estimate directly from the data. This observable representation is defined in the following low order moment matrices

$$[\vec{P}_1]_i = p(x_1 = i) \tag{2.30}$$

$$[P_{2,1}]_{ij} = p(x_2 = i, x_1 = j) \tag{2.31}$$

$$[P_{3,x,1}]_{ij} = p(x_3 = i, x_2 = x, x_1 = j) \forall x \in [n] \tag{2.32}$$

Probability vector $P_1 \in \mathbb{R}^n$, covariance matrix $P_{2,1} \in \mathbb{R}^{n*n}$, and 'trivariance' $P_{3,x,1} \in \mathbb{R}^{n*n}$, where each variable has a corresponding covariance structure of surrounding observations given that variable is observed. One further dependence is the matrix U where $U^T O$ is invertible, which is calculated through the SVD of $P_{2,1}$. The parameters of the dynamic model derived from these estimates are

$$\vec{b}_1 = U^T \vec{P}_1 \tag{2.33}$$

$$\vec{b}_\infty = (P_{2,1}^T U)^+ \vec{P}_1 \tag{2.34}$$

$$B_x = (U^T P_{3,x,1})(U^T P_{2,1})^+ \forall x \in [n] \tag{2.35}$$

P^+ refers to the Moore-Penrose psuedo-inverse of that matrix. All parameters of this reduced rank HMM have dimensionality in k, not m. Initial and normalization vectors (2.33, 2.34) can be related back to the original parameters, and in particular

$$B_x = (U^T O R) W_x (U^T O R)^{-1} \tag{2.36}$$

Making B_x a similarity transform of the reduced rank parameter W_x, which was the factorization of our original operator. These can be used to update estimates, from new data or through simulation

$$\hat{p}(x_1, ...x_t) = \hat{b}_\infty^T \hat{B}_{x_t} ... \hat{B}_{x_1} \hat{b}_1 \tag{2.37}$$

$$\hat{b}_{t+1} = \frac{\hat{B}_{x_t} \hat{b}_t}{\hat{b}_\infty^T \hat{B}_{x_t} \hat{b}_t} \tag{2.38}$$

$$\hat{p}(x_t|x_{1:t-1}) = \frac{\hat{b}_\infty^T \hat{B}_{x_t} \hat{b}_t}{\sum_x \hat{b}_\infty^T \hat{B}_{x_t} \hat{b}_t} \tag{2.39}$$

In Hsu et al, the initial proposal of spectral learning in HMMs was shown to be free of local optima and statistically consistent with finite-sample bounds on the error in joint probability estimates. In (Hsu et al., 2012) they show that many applications may have sample complexities independent of n, or regions of the brain, and the bound will depend polynomially on the m^th largest singular value of several of the matrices used in analysis, as well as t and m. Siddiqi et al extend this model, largely theoretically rather than in changing the algorithm, to learning reduced-rank HMMs. Learning in a low rank space still allows modeling arbitrarily large state spaces in $O(Nk^2)$ time. This large implicit state space allows smooth state trajectories, while the low dimensionality allows efficient learning. This generalized HMM algorithm is also extended to multi-step observations, rather than single steps in time, as well as high-dimensional real valued variables using kernel density estimation. They generalize bounds from Hsu to the case of low-rank HMMs, showing the consistency of the learning algorithm without needing the high-dimensional latent parameters. The number of samples are also dependent on k, rather than m.

If $\sigma_k(M)$ is the k^th largest singular value of a matrix M, the sample complexity depends polynomially on $\frac{1}{\sigma_k}(P_{2,1})$ and $\frac{1}{\sigma_k}(OR)$. As $\sigma_k(P_{2,1})$ increases, the better separated underlying dynamics are from noise. The term $n_o(\epsilon)$ is the minimum number of observations accounting for $(1-\epsilon)$ of the probability mass. They introduce this theorem: There exists a constant $C > 0$ where the following holds. Pick any 0 $\leq \epsilon$, $\eta \leq 1$ and $t \geq 1$. $\vec{\pi} > 0$, rank(T) = k, rank(U^TOR) $> k$ and rank(O) $> k$. Also assume rank($Sdiag(\vec{\pi})O^T$) $\geq k, ||R||_1 \leq 1$, and for some col-

umn c of R, $||R_c||_2 \leq \sqrt{\frac{k}{m}}$. $\varepsilon = \sigma_k(OR)\sigma_k(P_{2,1})\epsilon/(4t\sqrt{k})$. Assume $N \geq C * \frac{t^2}{\epsilon^2}(\frac{k}{\sigma_k(OR)^2\sigma_k(P_{2,1})^4} + \frac{k*n_0(\varepsilon)}{\sigma_k(OR)^2\sigma_k(P_{2,1})^2})log\frac{1}{\eta}$ With probability $\geq 1-\eta$, the model satisfies

$$\sum_{x_1,...,x_t} |p(x_1,...,x_t) - \hat{p}(x_1,...,x_t)| \leq \epsilon \tag{2.40}$$

Hsu et al mention that in practice, knowing the exact number of states is not essential as the spectral method tolerates models that are not exactly HMMs. Siddiqi et al go on to show that observable representations of reduced rank HMMs are a subset of k-dimensional Predictive State Representations, which are themselves a larger class of dynamic models that could be seen to subsume HMMs. This suggests that the spectral approach is a statistically consistent PSR learning algorithm as well, and it can apply to situations where the underlying model is not exactly an HMM, but still recover some sensible dynamics.

2.8 Implementation details

Example functions for spectral learning of dynamical systems were made available, implemented in Matlab code, at Geoff Gordon's website (http://www.cs.cmu.edu/ ggordon/specds/), and were used for applications in this work. This largely follows the steps as described above, the major computational step is in the singular value decomposition of the covariance structure, which is efficiently implemented in Matlab. Easily the largest use of memory is in calculating the trivariance structure $P_{3,x,1}$. To limit this usage, Gordon et al modify the algorithm in order to skip explicit calculation. The future-past covariance structure, which is effectively a cross correlation between offset time-series, $P_{2,1}$ is calculated, and a decomposition into U is done as outlined. These are used to create a basis, $U * P_{2,1}$ that is used for regressing

observations into their predicted states. This basis is also used to create future states by shifting sequences one time point into the future, accounting for the additional shift in $P_{3,x,1}$. The final B_x operator (2.35) is obtained through iterating over observations, weighing future states by current observations, and then taking the matrix product of this and predicted states. Finally, observations are multiplied with predicted states to create expected covariances of observations across states, $\hat{\Sigma}_{i,j} = p(x_{1,t} = i, x_{2,t} = j, b_t = s) \forall s \in [k]$, which are useful representations to help understand what dynamics are changing between transitions.

As Siddiqi et al discuss, it is possible to contruct Hankel matrices, concatenated time-shifted observations, which can be used to assess correlations between more distant time-shifted data in the past or future while maintaining the same system complexity k. In practice I did not see obvious benefits in the model to doing so, and covariance reconstruction tended to decrease with increased windows, though the effects of using different temporal distances wasn't investigated thoroughly.

These learned parameters can be used for further updates, as in simulations of time series. An update to expected observations is done through the tensor product of the modeled expectation and the prior state, and a state update is done through the tensor product of the observable operator and prior state. The observation probability is the diagonal of the updated observation matrix, and I use this as a surrogate for the observation in simulations.

$$\hat{O}_t = \Sigma \otimes \hat{b}_{t-1} \tag{2.41}$$

$$\hat{b}_t = (B_x \otimes \hat{b}_{t-1}) * (\frac{\hat{O}_t}{\vec{x}_t}) \tag{2.42}$$

$$\vec{x}_t = diag(\hat{O}_t) \tag{2.43}$$

These various expectations are not exactly probabilities, but are linear transformations away from HMM probabilities. This can lead to negative expectations, which in the original formulation would be in error. Siddiqi et al mention thresholding or modifying values that may fall too low. This topic was central to a recent critical view of spectral learning (Zhao & Poupart, 2014). They identify that spectral learning is an exciting method for various reasons, but they suggest negative probabilities have been ignored, and also mention that while this is an exciting technique EM is still successful and can outperform this newer method in some cases. I do often have negative probabilities which I normalize after updates, such as in (2.41-2.43). As I'll discuss in Chapter 4, while this is a general concern there was still a significant reproduction of results across distinct datasets, suggesting that while some of the bounds on error may be loose given our sample sizes, the underlying dynamics are still being recovered.

Functional organization and development in typical and autistic subjects

3.1 Introduction

Individuals with Autism Spectrum Disorder (ASD) have difficulties processing and reacting to emotional and social signals in faces. Typical brain development may involve changes to connectivity, particularly long-range connections between distal regions of the brain. These deficits could reflect a fundamental problem in face perception, and may be characterized by abnormal development of the brain, such as long range underconnectivity and local overconnectivity. A systems-level approach, like network analysis or using graph theory, may be well suited to assessing disrupted connectivity and in characterizing degree of functional integration and segregation. Greater segregation of functional networks in ASD could be an index of local over connectivity or more distant under connectivity. In this study I compared functional connectivity measured with fMRI during rest, face, and object viewing in children with ASD and matched neurotypical controls. I sought to investigate and compare functional connectivity patterns during rest and face processing to assess possible network differences, and to use a support vector machine approach to predict ASD status given those patterns. Additionally, it has been shown that these kinds of functional connectivity patterns can successfully predict age across typical subjects. I investigated whether those on the ASD spectrum could also match a similar maturation index, or if that index were disrupted in these subjects. I apply these techniques to a small group of children collected within the lab,

as well as to a much larger database of subjects provided publicly through a data sharing project.

This work was largely the focus of my early dissertation research, though results were mixed. Early analysis was on a limited sample, and after extending this work to the much larger dataset the picture actually became less clear. Nonetheless this is a useful opportunity to outline many of the steps involved in preparing and analyzing functional connectivity, and questioning why classification may have not worked well across these samples.

3.2 Methods

Participants Involved in Face Localizer Scans

Assent and consent were provided according to procedures approved by the University of Kentucky Institutional Review Board. Typically developing children were recruited through outreach presentations, fliers, email advertisements, and word-of-mouth. Children with ASD and siblings of children with ASD were recruited through the University of Kentucky Psychiatry Department and regional autism groups. The ASD-matched control and sibling subjects had and no first-order relatives with an ASD diagnosis.

To confirm ASD diagnosis or neurotypical status, all ASD, sibling, and matched control children were tested using the Autism Diagnostic Observation Schedule (ADOS; Lord et al. 2007). Parents completed the Social Responsiveness Scale (SRS, Constantino et al. 2003) and the Autism Diagnostic Interview âĂŞ Revised (ADIR, Rutter et al. 2005). The ADOS and ADIR were videotaped and results were confirmed via reliability testing by a speech-language pathologist. In addition, the Peabody Picture Vocabulary Test and Wechsler Block Design and Vocabulary

subtests were administered.

A typically developing control group of 23 children (all right-handed except 1 ambidextrous subject) were recruited as part of an ongoing study. 18 children with a prior ASD diagnosis volunteered for the study and participated in at least one visit. Five participants did not complete all visits and data from three participants were not usable due to excessive head motion in all functional runs. The final ASD group consisted of 10 males (7-17.8 years of age; M =12.86, SD = 3.31; 1 left-handed). ASD participants were asked to stop any medications on the day of functional scanning. 12 typically developing matched control children were originally recruited, data from two subjects were removed due to excessive head motion, leaving 10 control children matched by age, sex, and handedness (1 ambidextrous subject) to the ASD participants. Nine children (8 females) with a sibling with ASD volunteered for the study and participated in at least one visit. Sibling typically developing matched control group. 10 typically developing matched sibling controls were originally recruited, one participant was not neurotypical according to the ADOS/ADIR, leaving a final group of 9 control children (7 females) matched by age and handedness to the sibling participants.

Fifty-nine healthy right-hand adult volunteers (Mean age = 26.5, SD =6.0, range 18-42 years; 29 men) were compensated for participation or received course credit in an introductory psychology course. Due to excessive head motion (>1.75mm), data from eight participants were eliminated from analysis. No participants reported neurological or psychiatric diagnoses or pregnancy. All procedures were approved by the University's Institutional Review Board. Participants provided informed consent before participating.

Photo-realistic faces were constructed using FACES 4.0 software (IQ Biometrix, Redwood Shores, CA) and house stimuli were created using Chief Architect 10.06a

(Coeur dAlene, ID). The design was one where a preliminary 'face localizer' scan was run, where subjects passively viewed various face, non-face object, and noise stimuli, with interval blocks of no stimulus - a fixation cross at the center of the display. Subsequent designs used these as a basis for investigation of manipulations of these stimuli in various respects, but only the localizer scan was used here. All MR images were acquired using a Siemens 3T Trio MRI system (Siemens Medical Solutions, Erlangen, Germany). Subjects underwent one 109-volume (272.5 s) face localizer scan and four 133-volume (322.5 s) whole-brain functional task runs (pulse sequence: gradient echo, echo planar imaging; TE = 30 ms, TR = 2500 ms, flip angle = 80$12^\circ$, 64 x 64 matrix, FOV = 22.4 cm x 22.4 cm), interleaved acquisition of 38 transaxial contiguous 3.5-mm slices for the face localizer scan and 40 slices for the task scans. Following the functional scans, subjects underwent a high-resolution T1-weighted anatomical scan using magnetization-prepared rapid acquisition with gradient echo (MPRAGE) sequence (TE = 2.56 ms,TR = 1690 ms, TI = 1100 ms, FOV = 25.6 cm x 22.4 cm, flip angle = 12$12^\circ$, matrix size = 256 x 224) with 176 contiguous sagittal slices and a slice thickness of 1 mm. Field map information was also collected to correct geometric distortions caused by static-field inhomogeneity. The presentation of visual stimuli was controlled by E-prime software (version 1, www.pstnet.com; Psychology Software Tools) running on a Windows computer connected to the MR scanner. The time of each pulse of MR, each visual stimulus onset, and behavioral data responses were recorded by this software and used for data analyses. Preprocessing and statistical analysis were conducted using FSL (version 4.1.7, FMRIB, Oxford University, Oxford, U.K.) For each subject, preprocessing included motion correction with MCFLIRT and brain extraction using BET. The images were then smoothed with a 7-mm (FWHM) Gaussian kernel and were temporally high-pass filtered with a cutoff period of 100

s.

Additional subjects from ABIDE

Additional subjects were used from the publicly available Autism Brain Imaging Data Exchange (ABIDE) database which contains both ASD and age-matched control subjects undergoing resting-state scans across a variety of institutions (Di Martino et al., 2014). These institutions all follow their own scanning protocols and have multiple scanner types, and so their website should be consulted for details http://fcon_1000.projects.nitrc.org/indi/abide/. After downloading this data, field maps were used to correct for geometric distortions, images were motion corrected and spatially smoothed with a gaussian kernel (7 mm), then band-pass filtered between 0.01 and 1 hz.

Data preparation and Scrubbing

In each of 90 cerebral regions of the AAL atlas time series were extracted for the entire functional run and scrubbing of high displacement frames was applied as introduced by Power et al (Power et al., 2012). This measures a fractional displacement of images in the time series relative to the previous image, and summarizes the six rigid body motion parameters for movement and rotation already used for motion correction. Those images with corresponding displacement greater than 0.5 mm were eliminated from analysis in an effort to further reduce the effects of periods of high motion. In the case of face localizer scans, full time series was then broken into four separate time-series, one for each condition, faces, objects, textures and rest. Each of the experimental conditions' time series consisted of 21 time points and the rest condition consisted of 42 time points. Each time series

was shifted by one TR. Although time-shifting was not necessary, it more closely links the time points with the conditions of interest by accounting for the hemodynamic lag. Resting state scans from the ABIDE Dataset were analyzed as is after scrubbing.

Connectivity Matrices and Graph-theory metrics

For each subject and each condition, a connectivity matrix was computed from all pair-wise correlations across the 90 regions. Partial correlations were used to measure the degree of association between two regions time series while controlling for all other regions. This has the advantage of better reflecting the influence unique to those respective time series, and eliminates aspects of the signal that may otherwise dominate many regional correlations. The calculation is straightforward, but requires the inversion of the covariance estimate. In cases with the condition of âĂŸsmall n, large p', as our case where time points are significantly smaller than the 90 regions of interest, matrix inversion is not possible. To create a well conditioned covariance estimate, I apply a shrinkage factor as developed by Ledoit et al (Ledoit & Wolf, 2003) and implemented in Matlab by Shafer et al (Sch, 2005) . This finds a mixing parameter to solve a convex combination of the sample estimate against a diagonal target matrix.

Support Vector Machines

Support vector machines (SVM) are a common supervised learning algorithm for classification of labeled data. An SVM learns a representation of training examples in space, creating a map that separates label categories as widely as possible. This model can then be applied to additional data to evaluate their category.

Support vector regression is an extension to SVM that can be trained on real-valued outcomes such as age, as in dosenbach et al (Dosenbach et al., 2010), where an envelope of width epsilon is used to evaluate sample points along a regression line. To better interpret feature weights I use a linear classifier, which produced similar results to non-linear kernels across our dataset and produces weights which are more easily interpreted. All SVM and SVR calculations were performed with the Spider Matlab Machine Learning Toolbox. Leave-one-out cross validation (LOOCV) was used to ensure generality of the predictive models, where one unlabeled test subject in turn is predicted from a model trained on remaining labeled data.

Hierarchical Clustering

Group averaged matrices were submitted to a hierarchical clustering analysis (Matlab Stats toolbox) using ward's minimum variance method, which minimizes the total within-cluster variance in an agglomerative clustering assignment. This is a deterministic method that produces a tree structure relating groups of nodes to each other according to their shared distances. This tree can be cut to produce as many branches, or clusters, as desired.

3.3 Results

Using hierarchical clustering, subjects and groups largely have consistent patterns of organization. In Figure 3.1 I plot average functional maps of four groups, using rest blocks from the face localizer task. There is some re-organization across groups, the ASD subjects had a significant degree of lateralization (Figure 3.1a), whereas controls (Figure 3.1b) were generally bilateral. This lateralization was also

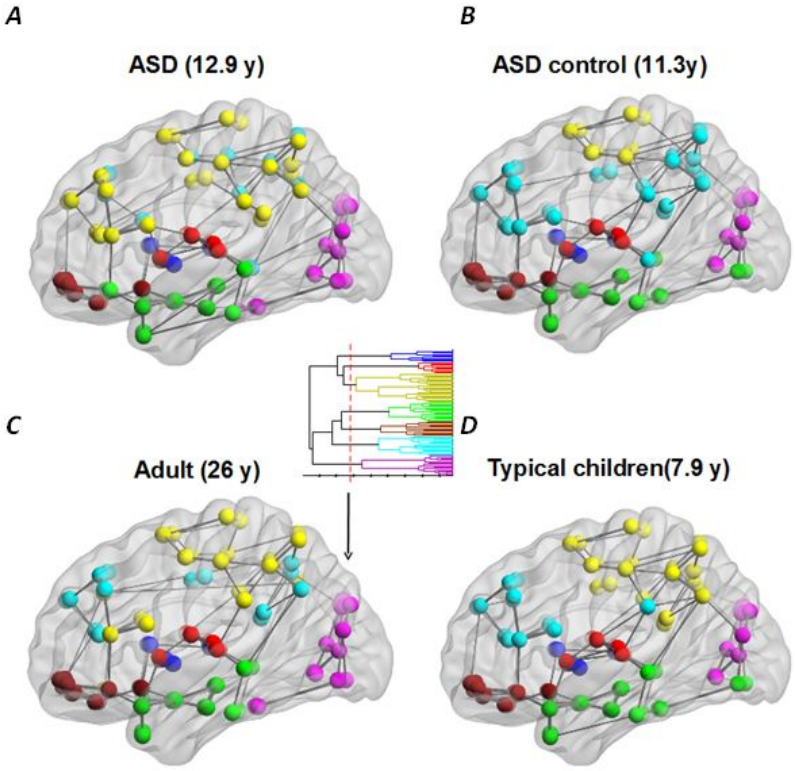


Figure 3.1: Average correlation matrices were used to generate clusters of regions functionally connected. Nodes are colored by cluster ID, assigned through hierarchical analysis, and the top 2 % of edges are plotted. Inset, in C, is an example dendrogram used for clustering.

seen in both typical children (Figure 3.1d) and adults (Figure 3.1c). There is some degree of continuation when looking at the young children, to somewhat older ASD controls, into adults, where fronto-parietal connectivity seems to be strengthening as a unique cluster. Inset I display an example of the hierarchy used to generate this figure, with the arbitrary cutoff that designates cluster assignments.

Applying the ASD and ASD control groups to a support vector machine classification yielded a high degree of classification accuracy, as seen in Figure 3.2. This was done through a leave-one-out approach that trains models, in turn, on all but one sample, and then tests on that subject. This model used the top 200 edge weights, as determined through a t-test between groups, for classification. To

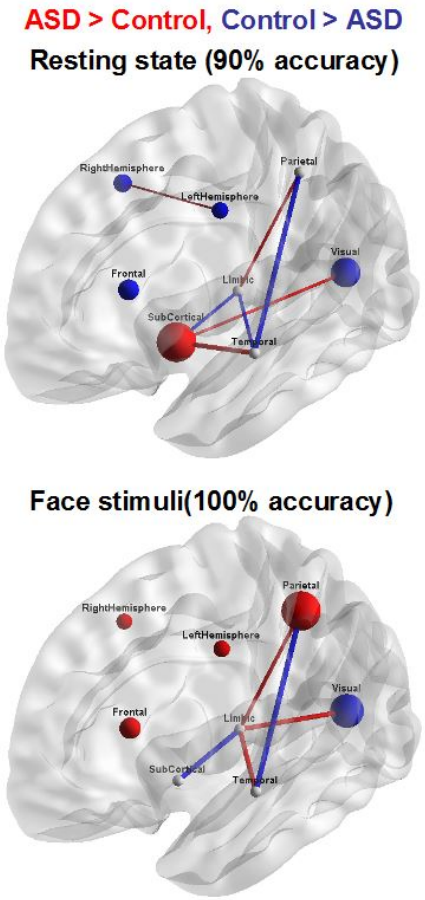


Figure 3.2: The top 200 edges discriminating between ASD and controls (via t-test) were used to create an SVM classifier for matrices generated from rest and face blocks. Brain network figures have had nodes and edges collapsed into respective regions, with red signifying more weight assigned to ASD subjects, and blue for controls. Node sizes reflect weight assigned to intra-node connections.

simplify their display I group them into larger regional interactions, for instance all frontal cortex edges that project to limbic structures are collapsed into one edge, and all edges between frontal regions are depicted through the size of that node. There are large differences between edge assignments in the two conditions, where red indicates areas where edges predicted autism, and blue indicates regions predictive of controls.

In Figure 3.3 I also applied a support vector regression analysis to typical sub-

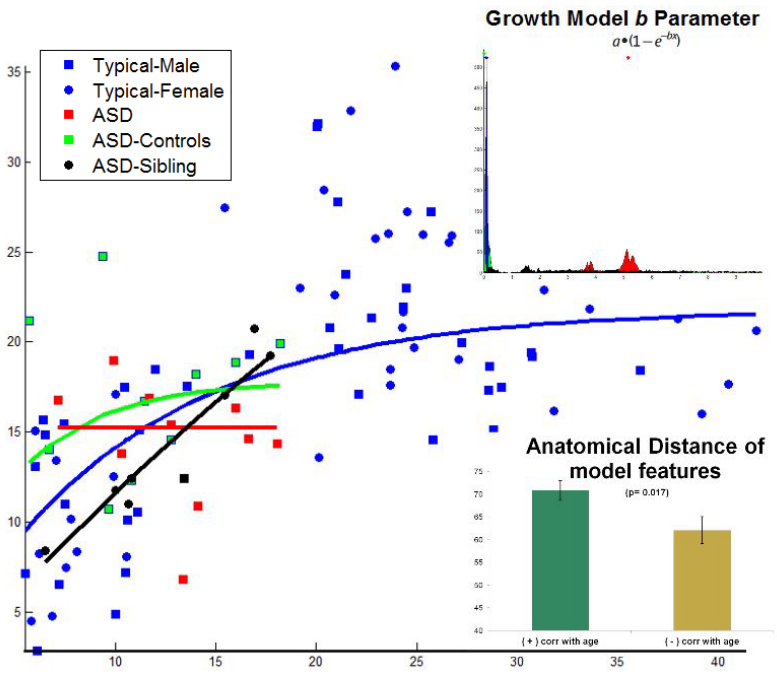


Figure 3.3: The top 200 edges, determined through individualy evaluating edge weight correlations with age, from task-negative correlation matrices were used as features to train an SVR on typical subject ages, in blue. This model was applied to the same data from the Autistic (red), ASD-control (green) and asd-sibling (black) subjects, and a growth model was fit to each sample.

jects (in blue), again with a leave-one-out strategy, which successfully creates maturation index with a relatively well fitting growth curve, seemingly plateauing at around 20 years old. Inset on the bottom right is a panel that shows a significant difference between the anatomical distance of features (inter-regional correlations) positively and negatively correlated with age. Autistic (red), ASD-control (green) and asd-sibling (black) subjects had this learned model applied to them, and their predictions are plotted. Inset in the top right are bootstrap estimates of a trained growth parameter for best-fit lines, suggesting that the autistic subjects have a much different, possibly flat or even inverse, growth curve.

The sample size of the autistic subjects was too small for a confident appraisal of these results, and so I applied similar strategies to much larger cohorts of sub-

jects from the ABIDE data sharing initiative. Figures 3.4 and 3.5 depict information on subjects and age distributions across centers. When looking at resting state connectivity from across typical subjects in this database (Figure 3.6b) it is different than that of continuous rest from an internal dataset from the lab (Figure 3.6a) While differences exist, the overall pattern of connectivity is relatively similar, on the same order as differences in 3.1, suggesting that the kind of functional connectivity routinely studied is present across subjects.

Figure 3.7 shows an analysis of the variability across typical subjects in the ABIDE data. Investigation into the inter-regional correlations shows a specific pattern mostly centered on temporal and frontal regions, though extending into subcortical and parietal areas, among others. While this distribution might be suggestive of some kind of center or scanner specific variability, further analysis suggests that these are actually intrinsic regions of functional connectivity variability within the brain.

In Figure 3.8a is a representative example of an identical prediction process learned from typical subjects, and applied to those with ASD. In this case it is difficult to differentiate between curves. I redid the analysis again on a collection of subjects across centers in 3.8b, and again ASD subjects are difficult to differentiate. Similar analysis as in Figure 2 was applied to predict whether a given subject's connectivity matrix was in the ASD or control group, success rates for each institution tended to lie from 50-70 percent. Interpretation of the most significantly discriminative edges was difficult, as different groups analysis yielded a different edge weight distribution.

In Figure 3.9 I show a dimension reduction using the top 3 principal components across all ABIDE subjects. The effect of center dominates these dimensions. Despite this the functional connectivity within these groups does tend to average into

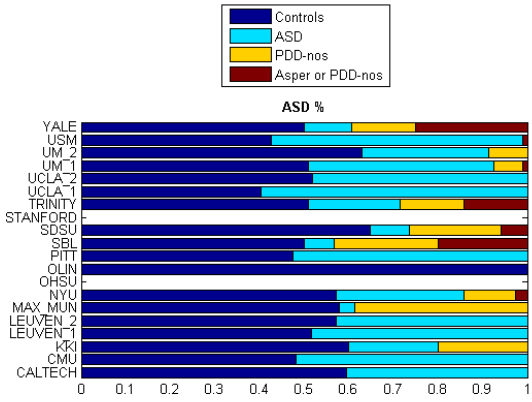


Figure 3.4: The distribution of control and asd subjects by institution within the ABIDE database.

familiar patterns. The NYU group in the lower left of this plot seems to be largely isolated, but as I'll show in the next chapter, models derived from that data are very similar to those made from internal lab datasets.

3.4 Discussion

The identification of biomarkers for autism would be a very useful translational tool for diagnosis and treatment in clinical decision making. A number of imaging techniques have been used to classify autism, both fMRI (Anderson et al., 2011; Wang et al., 2012) and other modalities (Ecker et al., 2010; Lange et al., 2010; Roberts et al., 2010). These approaches, all from single institutions much like our early results from the Kentucky sample, have approached 90% accuracy, with an average around 80%. Nonetheless these studies have not had consistent conclusions, and so the identification of biomarkers has been elusive despite the seeming success at classification.

Early work at identifying differences in ASD using fMRI found decreased functional activity (Just et al., 2004; Welchew et al., 2005), which was followed by many other reports of abnormal connectivity patterns (Belmonte et al., 2004; Rudie et

al., 2012), both during rest and under cognitive demands. Many of these were from decreases in connectivity in more distal regions, among a number of identified functional networks. While most reports were on decreases of connectivity, many have shown increases, or no changes, and the spatial location is largely inconsistent.

One of the principal concerns, and a likely reason for these general inconsistencies, was the typically low sample size used. Combined with the highly variable and still debated nature of autism itself, large sample sizes are likely needed to have the necessary power to detect possible biomarkers. The availability of a large multisite dataset from the ABIDE project was an important step in addressing this question.

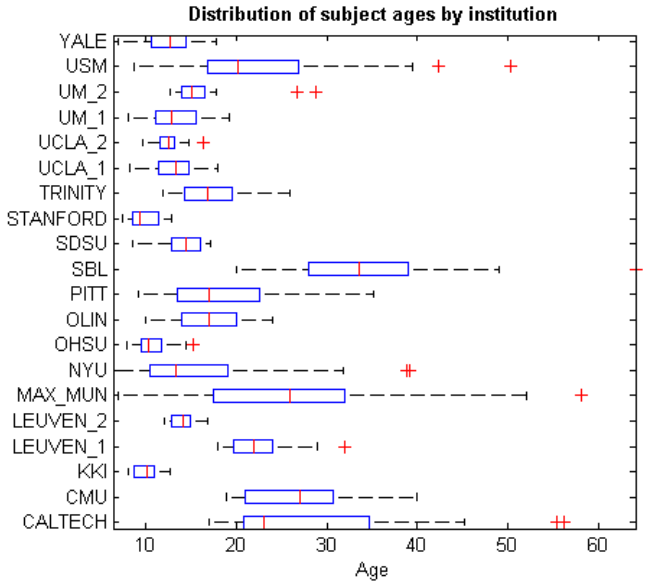


Figure 3.5: The distribution of ages in ABIDE by center

This dataset consisted of over 1000 subjects, split roughly evenly between typically developing and those on the ASD spectrum. This subjects were spread over nearly 20 institutions, and across numerous age ranges. There were many differ-

ent scanners, and a number of protocols involved in the various institutions, but the general design was a simple resting state scan. After comparing the average clustering behavior of networks in an independent continuous resting state dataset from the lab to a similar map generated across all typical subjects in ABIDE (Figure 3.6), it seems that while there are differences, generally speaking the organization is similar. This suggests that the sort of fundamental organization frequently described in literature is present across many subjects, regardless of scanners, protocols, and so on. This organization does change modestly when looking at individual centers in each institution, along the lines of the differences seen in this figure. There may be a general outline for brain organization into these modules, but no specific 'steady state', or alternately, inter-lab differences lead to modest modularity changes.

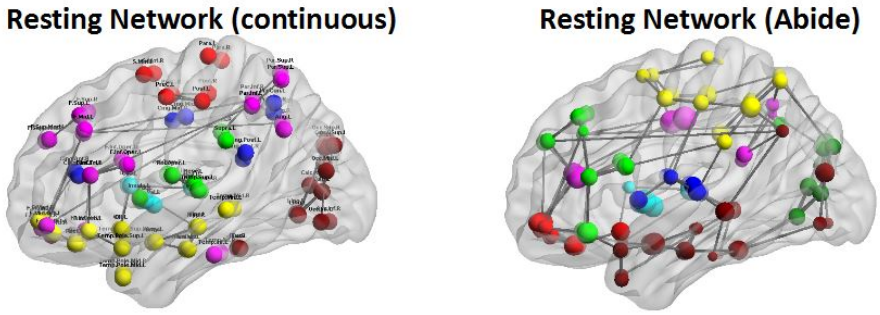


Figure 3.6: Average clustering generated from functional connectivity across subjects from an internal dataset (left), and all typical adult subjects in the ABIDE data (right). Nodes are colored according to cluster, and sized according to the sum of all edges into that region.

I also investigated the sources of variability in functional connectivity across all subjects. In Figure 3.7 I make a similar brain-map of this network of 'variability'. This indicates a network of regions largely organized around inferior frontal and temporal areas, but extending into subcortical, limbic, parietal, and occipital. While

this is largely an incidental observation here, these areas are also sources of variability when looking at within-subject functional connectivity, as I'll discuss in depth in Chapter 5. If the modularity of the brain is not in a stable state, there may be the appearance of some kind of variability mediating the shifts between states. This is an important concept going forward in this work, as the next chapter characterizes a picture of the brain that seems to always be shifting between various functional regimes, and Chapter 5 then suggests that this network of variability is another feature that may be coincident with those variable states.

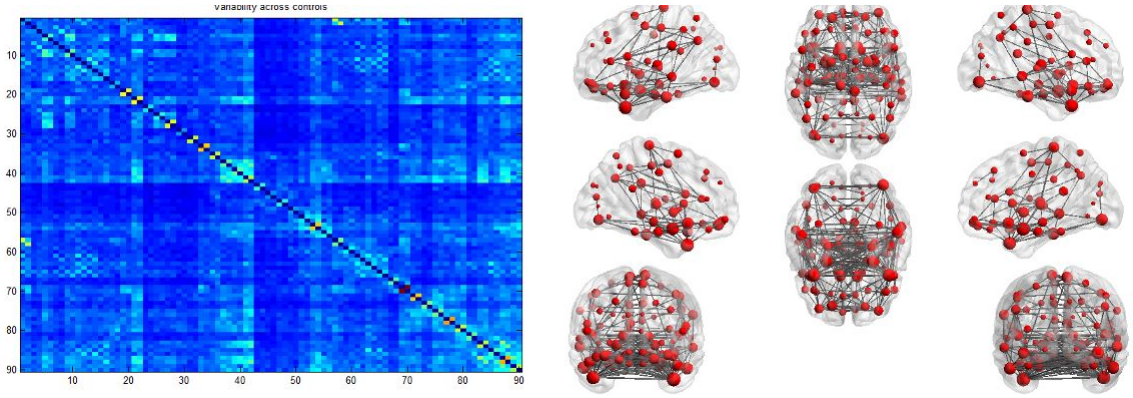


Figure 3.7: The variability of typical subjects functional connectivity across the entire ABIDE database is represented in matrix form (left) and various brain maps (right). It tends to organize in a fronto-temporal axis reaching into parietal and occipital regions. Despite significant differences in mean activation between centers (see Figure 3.9) variability in connectivity is reasonably consistent in these regions. This pattern of variability in the brains correlation structure persists in other datasets as well, and is the main topic in Chapter 5.

Finally I made a 'dimensionally reduced' representation of these subjects, as depicted in Figure 3.9. This shows that the greatest sources of variability is largely a function of center. Several other factors are modestly correlated with various eigenvectors, such as age or scanner type, but ASD's highest correlations are not significant, and are much lower than even the percent of data scrubbed (which is quite modest across the dataset).

The inconsistent nature of results found here were also found in another study looking at classification accuracy across this same data (Nielsen et al., 2013). There were some differences in their data preparation and methods, but the overall design and approach was similar - that is a multivariate classification using functional connectivity as features. They also obtained approximately 50-60 % accuracy across groups with generally different edge weights. I had also looked briefly at some other methods, such as logistic regression and deep belief networks (a type of artificial neural network) with similar conclusions. It seems that this particular approach of multivariate classification of ASD status using functional connectivity is unsuccessful.

The reason for this failure could be varied, including protocol differences, sampling from distinct populations, and scanner types. Many previous results did not specifically address rapid head movements, which I attempted to with scrubbing. This single factor may create disparities in many previous studies which often found potentially larger functional differences in distal regions than appropriate (Power et al., 2012). Motion has mainly been a concern in younger populations, but may also be a particular concern in atypical populations such as those with ASD. Of course, it is possible that while differences in individuals with ASD can be discerned in some fashion at the level of fMRI, the variability in the subtypes of ASD overwhelms detection of a specific marker. It may be possible for some of these classifiers to find edges that are consistent in the typical group, but more variable in ASD groups, yielding a moderate classification success while producing a difficult to interpret pattern of ASD activity. This goes hand-in-hand with the failure of analysis, where the appropriate 'features' to feed into these learning algorithms is still an open question. If this feature is quantifiable from fMRI it may not be a simple functional connectivity measure like correlation, or may be derived from it. Finally,

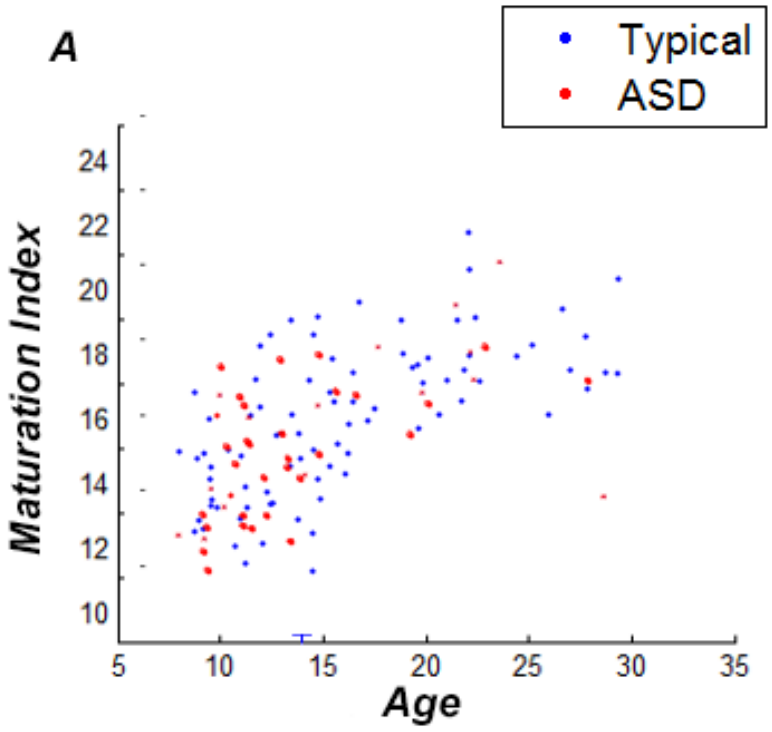


Figure 3.8: Prediction of subject age using continuous resting state data. A. A subset of the ABIDE data from NYU which had a good distribution across ages was used as in Figure 3.3. In this case after training on typical subjects (in blue), the autistic subjects (in red) are indistinguishable. B. A collection of subjects across centers was used to create a range of ages. Data were successfully used to create an index predicting subject age, though again the autistic subjects do not display any significant deviation from the growth model.

while SVM and other multivariate techniques had inconsistencies across groups, there may be other strategies available to use from the machine learning field that may be more successful.

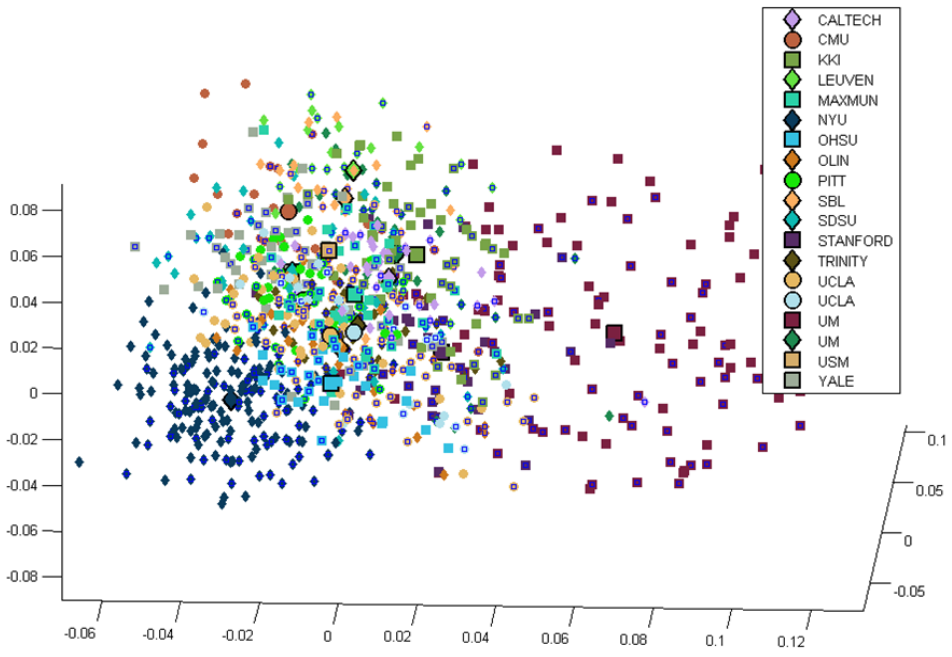


Figure 3.9: Principal components analysis of the top 3 dimensions from across all subjects in the ABIDE database. Some organizations have a modest mix, suggesting their subjects are similar, while others appear more isolated, suggesting there is some aspect to their data that separates them from other groups.

A Model of dynamic activity in the brain

4.1 Abstract

There is now considerable interest in the organization and development of spontaneous activity observed in the brain in the absence of any explicit stimulus or task. This brain activity, particularly observed in fluctuations in the blood oxygen level dependent (BOLD) signal in functional magnetic resonance imaging (fMRI), reliably organizes into sub-networks often resembling those observed during tasks. While this intrinsic functionality has been well characterized and often reproduced, the precise temporal dynamics of the brain during rest are not well understood and have only recently gained some attention. Here I measure how this organization changes over time as different subnetworks co-activate. I employ recent advances from machine learning to model the organization of different brain regions as their activity ebbs and flows over time. This approach, which is closely related to Hidden Markov Models, is applied to several distinct cohorts of subjects producing a number of reproducible patterns of activity, including but not limited to previously identified systems such as the default mode network. These results together suggest a dynamic brain with multiple interacting subsystems that are not discretely isolated elements but frequently overlap. These superpositions of modules provides insight into the dynamics seen not only during rest, but in tasks, where components of these networks tend to have small changes in activation levels, and I see this through the emergence of a ventral visual state during a passive viewing session.

Finally, I identify a possible way to simulate lesions and measure their effect on subsequent brain activity, potentially providing a method to probe pathological and disrupted brain states.

4.2 Introduction

While most fMRI studies of the brain focus on measuring changes in activity as a result of some task or stimulus, the importance of brain organization at rest is now widely accepted. The resting brain is metabolically active, with activity typically shifting during tasks less than 5% (Raichle et al., 2001) and statistically significant results are often much lower. Not only is the majority of brain metabolism active during rest, but brain organization itself, through a number of measures such as voxelwise analysis (Shulman et al., 1997) seed-based correlations (Fox, Corbetta, et al., 2006), data reduction techniques such as ICA or PCA (De Luca et al., 2006) and network measures like graph theory (Achard et al., 2006), has been characterized as having various components or subnetworks that arise during this spontaneous activity. The first and most studied of these was the default mode network (DMN) (Greicius et al., 2003; Raichle et al., 2001), a set of spatially separated cortical regions including medial prefrontal, posterior cingulate, precuneus, inferior parietal, and frequently parts of the medial temporal lobe. These regions seemed temporally synchronized and reduced in activation during tasks. Many other brain regions are also correlated during rest and this coherence seems to be modulated during tasks, rather than having functional organization generated purely as a result of some new context. These groups of correlated regions often include motor, dorsal and ventral visual networks, dorsal and ventral attentional networks, and the basal ganglia. Spontaneous activity has also been matched to behavioral

flexibility (Kelly et al., 2008), as well as shown to be developmentally dependent (Dosenbach et al., 2010; Fair et al., 2008). These findings have contributed to the growing interest in more explanatory models of brain organization above and beyond the traditional mean-signal changes (Hutchison et al., 2013; Smith, 2012), ultimately indicating the importance of better understanding temporal dynamics in and between functional regions of the brain.

The consistency of resting or spontaneous functional organization produces a picture of the brain that is characterized by a number of interacting sub-networks. This baseline, or homeostatic, state then reorganizes to reflect or accommodate various attentional or perceptual needs. Despite this consensus picture (Deco et al., 2013; Raichle, 2009), the actual dynamics of these observed systems and sub-networks have not been well characterized. While the existence of resting activity has been well established, is this activity static, in a single stable state waiting for some external perturbation, or is dynamic, possibly shifting between a number of different organized states that taken together average to the commonly characterized system? Evidence collected in the last several years has begun to identify the non-stationary nature of resting state activity, where the functional organization observed in many studies changes in magnitude over time. Chang et al show (Chang & Glover, 2010) that time-frequency coherence between posterior cingulate and anti-correlated regions was variable in both phase and coherence over time, and a sliding-window correlation revealed that a number of regions across the brain phased in and out of correlation with this region. The neural basis of a time varying signal was reinforced by finding that dynamic BOLD activity is often correlated to underlying local field potentials in rats (Thompson et al., 2013). Differences in dynamic default mode connectivity in subjects that positively correlated with a measure of mind-wandering suggest that not only is the DMN involved in introspective

focus, but the dynamics and frequency of coherence in this network may predict later behavior (Kucyi & Davis, 2014). Differences between controls and a group with Alzheimer's could be identified in the amount of time that specific resting state modules were maintained over time, rather than the static average (Jones et al., 2012). These findings all point to the importance of better characterizing dynamic brain connectivity.

The prevailing techniques traditionally used to inspect organization across the brain are fairly straightforward in application. One common approach is seed-based analysis, where correlations are made between one identified region and all other regions of the brain in an effort to find functionally âĂŸconnected' regions that share similar time-courses. Extensions of this approach to inspect all regions of the brain include independent components analysis and graph-theoretic analysis, which both look at all time-series simultaneously seeking functional groupings of spatially separated brain regions. These have all had success, are easy to implement, and require no explicit model building. Some other approaches attempt to better model dynamics of the influence of regions on connectivity, and have had more modest success and interpretation, these would include Granger causality or dynamic causal modeling (Smith et al., 2011). While these attempt to model directional or effective connectivity, they still show an 'average' influence between different regions across the entire observed time-series.

Other successful approaches in modeling complexity present in fMRI have been made through successful applications of tools from the machine learning literature. Golland et al use kmeans clustering to segregate the brain into two sub divisions, which they describe as intrinsic and extrinsic, integrated around DMN and sensory-motor respectively (Golland et al., 2008). More recently entire resting state networks were used with support vector learning to successfully predict

age in a developmental study (Dosenbach et al., 2010), and another study used spectral embeddings of brain structure to model patterns of dementia (Raj et al., 2012). Recently, several groups have begun to look at time-varying changes in the resting state signal. This approach may yield a magnitude greater degree of complexity as analysis is expanded into the time dimension, requiring special consideration as the number of parameters being tracked increases. Allen et al used a sliding window approach to cluster analysis (Allen et al., 2014) and find that functional connectivity does change over time, with some areas showing periodic patterns of higher degrees of cluster reorganization that may be missed in a static average. A similar sliding window approach was used to calculate principal components across time, and compared controls to a group with multiple sclerosis, yielding additional connectivity differences across time that were not seen in the time-averaged measure. (Leonardi et al., 2013) This study will integrate many of these recent approaches using a recently formalized approach from the machine learning literature that directly captures the dynamics as they change in time, and can then reconstruct, through simulation, the observed variability as needed.

Hidden Markov models (HMM) have been a very successful tool for understanding time-series dynamics (Baum & Petrie, 1966). Briefly, HMM's are the simplest dynamic Bayesian network, where the system being modeled is assumed to be a Markov process with some unobserved or latent states driving observations, typically discrete or Gaussian in nature. The state determines a probability distribution over the possible outputs, and so as this state changes over time the observed time series will consequently change. They have been used successfully in many applications, perhaps most recognized in speech, handwriting, and gesture software, as well as a variety of genomic and proteomic bioinformatics uses (Stigler et al., 2011), and have had some success in neuroscience related studies, though

most often in studies looking at ion channel or spike-train dynamics (Danóczy & Hahnloser, 2005; Herbst et al., 2008). The learning problem is then: given the output sequence such as an fMRI time series, how can a model optimally learn the unknown brain states that are driving the observations. Once a model of these hidden states is created, the learned representation can be evaluated against known patterns, such as the DMN, and their dynamics can be tested - do they remain largely constant, vary over time, or appear random. There is an inherent difficulty in solving the underlying parametric problem, and approaches have typically relied on traditional and well accepted methods such as maximum likelihood and expectation maximization (EM). Applications to fMRI data are rare and typically model task based data (Faisan et al., 2007). The first application of an HMM to resting state data was by Eavani et al who use an HMM framework with sparse basis learning to identify that functional connectivity over time can be modeled as a combination of different basis networks (Eavani et al., 2013). This analysis used a joint learning strategy that effectively generated prior basis to simplify and constrain traditional HMM learning. These basis were constrained to rank-one decompositions of covariance, as well as restricting an l_1 norm to reduce the number nodes in each basis set. While this was an important work, this approach may limit the ability to detect the underlying dynamics, as basis are evaluated from static covariance. Additionally the study only investigated cases where the number of states was equal to the number of basis used. While they showed good results in recovering a simple simulated network, it is unknown how well their state recovery performed in the real data. One challenging issue in modeling dynamics is the relationship between the number of systems in the brain, and the number of states the brain can exist in. Traditional HMM approaches, including those done by Eavani et al, typically have a 'rigid' state assignment, which imposes a discrete view of fluctuations between

possible organizations, when the brain is actually acting in a much more continuous fashion. A recently discovered approach to solving these dynamical systems may introduce a method that can address many of these concerns, while learning an underlying low rank basis set directly from temporal patterns in the data. This low rank approximation to an HMM allows one to model systems with a much higher state space, and uses a new approach to parameter learning using a class of spectral learning algorithms which avoids issues of local optima such as EM. These use a generalization of dynamical systems that include HMMs, Predictive State Representations (PSR) and the closely related class of algorithms known as observable operator models. (Boots et al., 2011; Hsu et al., 2012; Siddiqi et al., 2009; Singh et al., 2004)

Here I apply these spectral learning techniques, as implemented by Boots et al (Boots et al., 2009), to construct a dynamic model of the brain that can explain most of the observed variability. Through easily implemented linear algebra methods, namely a singular value decomposition of time-shifted covariance structures, this learning approach is statistically consistent, avoids issues of local optima, and has memory usage independent of the number of training examples, with time complexity linear to this number - most of the results in this study took seconds to run on an average PC. This method allows for very large data sets, and can scale well to much more complex models than many prior approaches. I show the ability to capture most of the dynamics associated with resting activity, identify consistent subnetworks across subjects and groups, and are able to reproduce the observed covariance structure of the data through simulation. These discovered dynamics appear to reflect a rapidly transitioning and superimposing background of brain activity, modulating between a number of systems including previously identified networks such as the DMN. I also investigate a group of subjects passively view-

ing visual stimuli, and while these subjects overall states remain largely consistent, there is a rearrangement of these states with the ventral visual stream becoming more pronounced. Finally I introduce an approach to investigating synthetic lesions in the learned model, which presents a new method to test disruptions in these complex systems.

4.3 Methods

fMRI data acquisition

Dataset 1 contains continuous resting state data from fifty-one healthy young adults (18-25 years), recruited under my advisor's grant award, from two different institutions (15 subjects were enrolled at the University of Kentucky and the remainder of the subjects were enrolled at the Medical University of South Carolina). The images in each participant's time series were motion-corrected, geometric distortion corrected, spatially smoothed with a 3-D Gaussian kernel (full width at half maximum = 7 mm) using FSL v. 4.1 (http://www.fmrib.ox.ac.uk/fsl). The statistic parametric maps were then registered via the subject's T1 anatomical scans to the MNI-2mm template. Data were bandpass filtered from 0.01 to 0.1 hz, as is common in resting state literature.

Dataset 2 included subjects collected from a single institution (NYU) from the publicly available ABIDE database which contains both autistic and age-matched control subjects undergoing resting-state scans, across a variety of institutions (N=82,mean age 16.1). I use all eye-open controls from this subgroup. Preprocessing was performed as above.

Dataset 3 contained 51 healthy right-hand volunteers (Mean age = 26.5, SD =6.0, range 18-42 years; 29 men) were compensated for participation or received

course credit in an introductory psychology course. The design was one where a preliminary 'face localizer' scan was run, where subjects passively viewed various face, non-face object, and noise stimuli, with interval blocks of no stimulus - a fixation cross at the center of the display. Subsequent designs used these as a basis for investigation of manipulations of these stimuli in various respects, but only the localizer scan was used here. All MR images were acquired using a Siemens 3T Trio MRI system (Siemens Medical Solutions, Erlangen, Germany). Subjects underwent one 109-volume (272.5 s) face localizer scan and four 133-volume (322.5 s) whole-brain functional task runs (pulse sequence: gradient echo, echo planar imaging; TE = 30 ms, TR = 2500 ms, flip angle = 80Âř, 64 x 64 matrix, FOV = 22.4 cm x 22.4 cm), interleaved acquisition of 38 transaxial contiguous 3.5-mm slices for the face localizer scan and 40 slices for the task scans. Following the functional scans, subjects underwent a high-resolution T1-weighted anatomical scan using magnetization-prepared rapid acquisition with gradient echo (MPRAGE) sequence (TE = 2.56 ms,TR = 1690 ms, TI = 1100 ms, FOV = 25.6 cm x 22.4 cm, flip angle = 12°, matrix size = 256 x 224) with 176 contiguous sagittal slices and a slice thickness of 1 mm. Field map information was also collected to correct geometric distortions caused by static-field inhomogeneity. The presentation of visual stimuli was controlled by E-prime software (version 1, www.pstnet.com; Psychology Software Tools) running on a Windows computer connected to the MR scanner. The time of each pulse of MR, each visual stimulus onset, and behavioral data responses were recorded by this software and used for data analyses. Preprocessing and statistical analysis were conducted using FSL (version 4.1.7, FMRIB, Oxford University, Oxford, U.K.) For each subject, preprocessing included motion correction with MCFLIRT and brain extraction using BET. The images were then smoothed with a 7-mm (FWHM) Gaussian kernel and were temporally high-pass

filtered with a cutoff period of 100 s.

Prior to parameter learning of dynamics, all time series data are normalized in a range 0-1.

Partial Correlations and Hierarchical Clustering

When looking at static time-averaged patterns of connectivity, partial correlations were used to measure the co-linearity between two nodes while controlling for all other nodes. These have been shown to be more sensitive to specific node to node connectivity as they reflect the more direct relationship between two time-series (Marrelec et al., 2006; Smith et al., 2011). If P is the inversion of the covariance matrix of the time series data, the partial correlation ρ between nodes i and j is: $\rho_{ij} = \frac{p_{ij}}{\sqrt{p_{ii}p_{jj}}}$ These partial correlation matrices were submitted to a hierarchical clustering analysis (Matlab Stats toolbox) using ward's minimum variance method, which minimizes the total within-cluster variance in an agglomerative clustering assignment. This is a deterministic method that produces a tree structure relating groups of nodes to each other according to their shared distances. This tree can be cut to produce as many branches, or clusters, as desired.

Parameter Learning

In contrast to traditional HMM's, which learn the transition probabilities between some unobserved latent states, as well as associated state dependent observation probabilities, I use a Predictive State Representation that creates a model of dynamic states by measuring the occurrence probabilities of future events conditioned on past events. Learning is done through a spectral algorithm, which can produce observable operators that are a transform of a reduced-rank HMM

(RRHMM), where dimensionality depends on rank rather than state space size. This allows a large implicit state space where the model can still have smooth state trajectories and a compact, efficient learning procedure.

This method and implementation were discussed in depth in Chapter 2, and more details can be found there. Briefly, this approach begins with definitions from the HMM literature, which use a transition matrix $T \in \mathbb{R}^{m*m}$ to identify probabilities of moving between states where the current state, of m possibilities, is understood to be tracked over time in z_t. An observation matrix $O \in \mathbb{R}^{n*m}$ identifies observation probabilities given the current state of each variable in vectors $\vec{x}_t$, for example our n regions of the brain. There is also an initial state probability vector $\vec{\pi}$, where $\pi_i = p(z_1 = i)$. The alternative view of computing sequence probability from observation operators used in this work defines

$$A_x = Tdiag(O_{x,1}, ..., O_{x,m}) \tag{4.1}$$

where for any time t

$$p(\vec{x}_1, ..., \vec{x}_t) = 1_m^T A_{x_t} ... A_{x_1} \pi \tag{4.2}$$

A_x is a tensor $\in \mathbb{R}^{m*m}$ $\forall x \in [n]$ and diag(O) is a diagonal matrix with the respective observation probabilities of a variable under a given state. This assumption can be relaxed, where the underlying system as modeled in T may have low rank, k, where $T = RS$, $R \in \mathbb{R}^{m*k}$ and $S \in \mathbb{R}^{k*m}$. This casts the initial state distribution into a lower dimensional state, $\pi = R\pi_l$, and likewise A can be projected into this space using a related matrix W:

$$p(\vec{x}_1, ..., \vec{x}_t) = 1_m^T R W_{x_t} ... W_{x_1} \pi_l \tag{4.3}$$

With this formulation, I can avoid using EM and estimate a linear transform of W directly from the data. This observable representation is defined in the following low order moment matrices in 3 general steps. First, I compute empirical covariance estimates $\hat{P}_1$, $\hat{P}_{2,1}$, $\hat{P}_{3,x,1}$ directly from the data, which respectively describe initial probability vectors of the observed data, the covariance of time-shifted time series, and future-past covariance between regional time series as a function of each specific region. Then I use a singular value decomposition of $\hat{P}_{2,1}$ to compute $\hat{U}$, the matrix of left singular vectors for each of k largest singular values from $\hat{P}_{2,1}$. From these empirical estimates I compute the model parameters $\hat{b}_1$, $\hat{b}_\infty$, $\hat{B}_x$, which relate back to the original model parameters, for instance

$$B_x = (U^T OR) W_x (U^T OR)^{-1} \tag{4.4}$$

This makes B_x a similarity transform of the reduced rank parameter W_x, which was the factorization of our original operator. These are used as model parameters, and can be used to update estimates, from new data or through simulation.

This algorithm is idealized for sampling N triplet sequences from a dynamical system. In practice this is more often 1 or more long time-series which can also be used to develop estimates, and this extension is also detailed by Siddiqi et al. Example functions for spectral learning of dynamical systems were made available, implemented in Matlab code, at Geoff Gordon's website http://www.cs.cmu.edu/~ggordon/specds/, and were used for applications in this work. This largely follows the steps as described above, the major computational step is in the singular value decomposition of the covariance structure which is efficiently implemented in Matlab. Easily the largest use of memory is in calculating the trivariance structure $P_{3,x,1}$. To limit this usage, the algorithm was modified in order to skip explicit

calculation. The future-past covariance structure, which is effectively a cross correlation between offset time-series, $P_{2,1}$ is calculated, and its decomposition into U is done as outlined. These are used to create a basis, $U * P_{2,1}$ that is used for regressing observations into their predicted states. This basis is also used to create future states by shifting sequences one time point into the future, accounting for the additional shift in $P_{3,x,1}$. The final B_x operator is obtained through iterating over observations, weighing future states by current observations, and then taking the matrix product of this and predicted states. Finally, observations are multiplied with predicted states to create expected covariances of observations across states, $\hat{\Sigma}_{i,j} = p(x_{1,t} = i, x_{2,t} = j, b_t = s) \forall s \in [k]$, which are useful representations to help understand what dynamics are changing between transitions.

Simulations and lesions

For simulation of timeseries from learned models, an update to expected observations is done through the tensor product of the modeled expectation and the prior state, and a state update is done through the tensor product of the observable operator and prior state. The observation probability is the diagonal of the updated observation matrix, and I use this as a surrogate for the observation in simulations.

$$\hat{O}_t = \Sigma \otimes \hat{b}_{t-1} \tag{4.5}$$

$$\hat{b}_t = (B_x \otimes \hat{b}_{t-1}) * (\frac{\hat{O}_t}{\vec{x}_t}) \tag{4.6}$$

$$\vec{x}_t = diag(\hat{O}_t) \tag{4.7}$$

The x vector is normalized from the range 0-1 at each update. For visualization I apply the same low pass filter, from 0.01 to 0.1 hz, that was applied to observed

data. For synthetic lesions, I either artificially set a given regions value in x to zero, which influences further state updates, or I set all respective entries in the expected observation matrix O_t to zero to eliminate the influence of a given state.

Visualization of anatomical brain maps and nodes was done using BrainNet Viewer (Xia et al., 2013), transition graphs were made with Gephi software (*Gephi: an open source software for exploring and manipulating networks*, n.d.).

4.4 Results

Brain organization is observed in the correlation structure of fMRI time series

Using static correlation matrices generated from subject time series data (Figure 4.1A), I reproduce graph theoretic results as often seen in the literature (Rubinov & Sporns, 2010; Salvador et al., 2005). This can be used for further analysis, such as one identifying modules or subnetworks (Figure 4.1B). These yield many common observations, such as strong bilateral connectivity and grouping of common functional systems including visual, fronto-parietal, sub-cortical and motor. These static networks do not always clearly reproduce results found from other analysis, that is, the default mode network is most clearly identified through seed based approaches, and often through ICA, but may appear grouped with other regions in clustering analysis.

Decoding functional dynamics in a single subject

Rather than using the simple average covariance of the signal to analyze functional connectivity, I can generate a covariance between past and future observations, and use a spectral approach to identify the top dimensions (through a

singular value decomposition) that is used as a basis for understanding the system. Regression using this basis can identify amplitude of states, where each state corresponds to that identified singular vector. In Figure 4.1C two of the top state amplitudes or weights are plotted in time, with Figure 4.1D showing their respective expected observations. Only two of the 12 modeled states are shown. In D, the observations are split into positive and negative regions (red and blue). A consequence of the singular value approach is that the sign in both the state amplitude and its respective expected observation are arbitrary, and these parameters may in fact change given small changes in the data. The actual observation can be understood as the weighted combination or superposition of these expected observations by their state in time, and so as the sign and magnitude of the state vector changes in time so does the output of the system.

This case in a single subject is illustrative of two important questions being asked. First is whether the system is largely stationary, and clearly there is a large flux in how the brain is organized over time. In Figure 4.1C I plot the amplitude of two components of the state vector over time. These are effectively the weights of an associated basis shown in 4.1D. These weights fluctuate considerably from positive to negative, and at times they are highly synchronized while at other times are nearly anti-correlated. This is characteristic of learned state vectors across subjects, and there were no cases where a subject's functional organization was stationary over time for long durations, or which appeared to be mainly composed of only one or two states. A second question is towards the nature of the brain states and how they are organized over time. There are a number of subsystems in the brain that are regularly learned in single subjects and across groups, including the two plotted in Figure 4.1. State 2 in particular is closely related to the commonly studied default mode network, plotted above in red in D. This incorpo-

rates medial frontal, posterior cingulate, and inferior temporal/angular gyrus. In C these two states start out largely synchronized, but then have large periods of anti-correlation. This reflects an important finding of the study, that while these base states are consistently learned across subjects and different centers, they super-impose to create additional intermediate states. This is simultaneously a strength and weakness of this approach. While more traditional Hidden Markov models typically simplify assumptions to model a system as being in a single discrete state, this approach produces a state vector that may better reflect that complex organization of the brain, which is by no means guided by unitary modes across all regions. The trade off is the loss of explicit modeling of the transition probabilities between states and over time, though I can recover the observed operators that are transforms away from being related to these factors. In 4.1E I plot a graph (left) of the mean observed transition matrix across observables or regions. This is a reflection of the influence of observations, which are marginalized out by the mean, to the next state in the system. On the right is the matrix generating the graph, with rows and columns relating to states 2-7 of the modeled system. The diagonal reflects a weight for remaining in the same state, and is reflective of the approximately 2-4 time steps that the system seems to stay in a given functional mode above. There is some considerable variance across subjects in these observed transition parameters, and this likely reflects the unconstrained nature of the resting state systems. Regardless of instructions to the subjects, there is likely a large degree of subject variability in how the mind may wander, attend to stimuli, and so on, which could then influence the change in brain sub-system activity.

The dimensionality of the model is largely consistent across subjects and groups

An important and useful aspect of the spectral approach to learning is that there is a natural relationship between the order of the states and their explained variance, similar to the proportionality of singular values or eigenvalues to variance in principal components analysis. This can be used to guide the number of states used in analysis. In Figure 4.2A each subject in Dataset 1 has the 2nd-12th singular value plotted, over the sum of all singular values, with the singular values resulting from the full model across all subjects in bold. The first singular value was not plotted as it tends to dominate in magnitude. Most subjects follow a similar trend, with a gradual plateau down with increased dimension, though a few subjects appear as outliers that can be modeled by fewer states. This Figure can be compared to Figure 4.6, where I compare simulated time series, using variable number of states, to the observed group average. Simulations using 3-25 states are generated, and each corresponding simulated correlation matrix was itself correlated with the observed matrix, and the R^2 increases similarly to the change in singular values. This shows that as additional states are learned through the model, additional information is captured and reflected in the measured correlation structure. As more states are added they effectively prune connectivity from the mixture of state representations until increasingly accurate representations are produced. It should be noted that partial correlations were used in lieu of traditional Pearson correlations, which are distinct in that they capture direct relations between time series conditioned on other time series, which is not something explicitly captured in the model. There is a generally linear increase in variation explained as states are added to the model, until the explained variance plateaus in the teens.

In Figure 4.2b I compare the full model, generated across all subjects, to individual models produced in order to better understand how consistent state discovery is across subjects. The magnitude of the correlation between the expected observation of state two in the full model, above, and seven, below, are indicated against each subjects individual states. This shows that the second state, the putative DMN (table 1), is generally consistent appearing in the top 2 or 3 states within most subjects. State 7 is less precise, and is not always as clearly correlated with the full model.

The identification of a base set of functional subnetworks

For further analysis I compare models learned across subjects within each respective dataset. Dataset 1 and 2 were recorded at different institutions, on different scanners, but were both from the same design, namely eyes-open continuous resting scans. Both appear to find very similar basis used to determine the dynamics in the system. In Figure 4.2C the top 6 states expected observation from Dataset 1 are plotted in anatomical space. The color indicates polarity, as shown in Figure 4.1, any state can have a weight and sign, this modifies the observation and so a negative state would flip the polarity of the observation. The size is proportional to the magnitude of the expected observation, and I also plot the top 2 percent of the underlying expected covariance as edges between regions. In all cases State 1 is a pseudo-constant state that accounts for a large percent of the observed signal. This seems to be largely related to the global signal, or a global basis. State two is the basis containing the default mode network, including posterior cingulate and medial frontal regions, but wrapping around frontal and temporal areas. The opposing signal comes mainly from motor and sensory areas. The third state

has a large component centered on visual regions, from occipital to the fusiform gyrus, and opposite nodes from insula, supramarginal gyrus, palidum, and putamen. The fourth had a strong bilateral fronto-parietal network, with opposition from a limbic system mainly surrounding hippocampus. State 5 had a network connecting inferior frontal with middle temporal and the temporal pole, with an opposite network spanning subcortical regions to posterior cingulae and parietal. State 6 had strongly lateralized networks, one being a more left dorsal fronto-parietal, the other a more ventral fronto-parietal. These may correspond to attentional networks, particularly the more ventral right lateralized network. The seventh state (Figure 4.3 for state 7-10) has a somewhat left-lateralized cluster bridging inferior occipital, inferior temporal, into frontal regions, with an opposite network spanning more dorsally from medial frontal through the cingulate. The former seems likely to correspond to the ventral visual stream, and as shown later does appear to be important in the visual task related group. State 8 has nearly spatially orthogonal networks, one bridging occipital and primary visual regions to anterior cingulated and frontal, the other spanning hippocampal and temporal to parietal. State 9 is moderately lateralized and one network spans the temporal pole, middle temporal, and medial frontal, while the other is mainly organized around inferior frontal and insula. State 10 has a network focused on motor areas, as well as possibly an auditory network mainly bridging Heschl's gyrus and the angular gyrus. State 11 has a moderately left lateralized network bridging superior temporal pole, cingulate, and parietal areas, and a bilateral network bridging caudate with pre and post central areas. Finally state 12 has two strongly focused networks focused on the gyrus rectus, and caudal areas, respectively. See table 1 for a detailed description of the most pronounced regions in each state.

The modeled basis driving resting dynamics is consistent between groups

The analysis done on Dataset 1 was reproduced in Dataset 2, and many of the states were closely reproduced. Figure 4.3b shows a cross-correlation matrix between respective expected state observations in each group. High values along the diagonal suggest a match between groups. State 1-4 shows a close match between groups, 5 is a modest match. States 6 and 7 match, but have reversed places, suggesting a slight different in proportion of variance explained in each. State 8 is a relatively poor match, while 9, 10, and 11 are quite good. As noted the matches are modest into the teens but it seems that these top 10-12 states account for most of the reproducibility between groups. States 5, 7, and 8 may appear to be 'leaking' somewhat between groups, with some degree of correlation between them. Figure 4.3a shows a number of the states matched between the two datasets, and the general similarities or lack thereof. The principal difference in state 8 seemed to be a shift from the ventral-dorsal connectivity between parietal and temporal regions, to a more frontally organized subsystem.

The utility of the approximate transition matrices that can be made could be evaluated by testing how related they are between groups. In Figure 4.3c the off-diagonal elements of the state transitions are plotted and correlated against each other, with a significant correlation between them. There is a higher correspondence between the transitions in the first few states, as shown in green. A graph of the transitions from Dataset 1 is shown in Figure 4.3d. The reproduction of some aspects of these transition observations suggest that there is some structure to state transitions at rest, and they are not purely random fluctuations between possible states.

A shift in the functional basis due to perceptual task

Having established that there is a consistent set of states driving resting states, I look at how these states may change during a simple passive viewing task. In Figure 4.4a, states in Dataset 1 are compared to those in Dataset 3, with the viewing task. In this case, there is clearly less correspondence between states compared to Dataset 2. This is largely due to shifting of states, particularly 3, 5, and 7. The 7th state in Dataset 1, which included a ventral network integrating visual, temporal, and frontal areas, becomes the third most important state driving these observations (behind the default mode network state). This identified state is shown in Figure 4.4b. I can also evaluate what states, if any, correlate with the identifiers in the design matrix (ie 1 for task, 0 for rest). The third state was most correlated($R^2 = .38$), more than double the next closest, state 4. In C, a single time series from a subject with a high correlation between these states and the design is shown, with the amplitudes for state 2 and 3 below, a potential index for when the brain is using a default mode, and when it is using the ventral visual network. Above the plots for states are blocks in red for visual task, and black for a blank screen. Those areas of each line that are the max for that time point are in bold, and a pattern emerges for the switching between brain states. In D the graph for the group transition matrix is plotted, and there may be a bit more bi-stability forming, as would be expected when inducing multiple states throughout the experiment, with a cluster of states around 8, 9, 10, and 11. There is no correlation between the off diagonal transition observations between this and Dataset 1, in contrast with the two resting state designs.

Simulated activity and lesions

A consequence of the formal modeling of the dynamics of fMRI activity is the ability to use that model to simulate time series data closely mimicking the original observations. In Figure 4.5a a single subject's timeseries from Dataset 1 is plotted with a longer simulated time series on the right. At time point 500, a synthetic 'lesion' of activity in the posterior cingulate is introduced, effectively reducing the activity in that node. It is difficult to perceive the change in dynamics during that time, but the relative change is shown in B, where correlations and regions which were higher prior to lesion are indicated in blue, and those that are higher during the lesion are in red. Assessing the assigned states over time, there are significant differences in the amplitude of states between these two periods in a number of states, particularly 2, 6, and 7 are all states heavily incorporating the posterior cingulate. Additionally, I briefly investigated the ability to remove a given state, in C results from the synthetic removal of the second, default mode, state are shown. Not surprisingly there is a reduction in the default mode, but there is a particular increase in a set of nodes oriented around the middle temporal pole. These investigations are preliminary but show additional potential of the method to investigating disruptions to brain networks.

4.5 Discussion

Spontaneous activity was first formalized by Raichle et al (Raichle et al., 2001), and was based on earlier meta-analysis (Shulman et al., 1997) as a default or baseline state of activity. The regions showing consistent "deactivation" , that is, increased activation during baseline compared to active tasks, included the posterior cingulate and precuneus, inferior parietal, and medial frontal regions. Further results

often describe connectivity between these and other regions through time series analysis, typically using correlations, reflecting some positive or negative co linearity between the respective signals. Since first reported, the DMN has been consistently observed, most consistently at low-pass filtered, continuous data (Cordes et al., 2001) similar to Datasets 1 and 2 from this study, but also from block design task-negative studies, such as the passive face viewing in Dataset 3. This network is often described as having a negative correlation with attention networks, with the two often being framed as task-negative and task-positive, suggesting a neural architecture mediating between periods of introspection and periods of attention to external factors and action. While some studies find hippocampal coactivation with other DMN units (Greicius et al., 2003), other studies are unable to reproduce this finding (Damoiseaux et al., 2006). This could reflect the non-stationary nature of these subnetworks, as some analysis may yield a result reflecting the occasional superposition of these states. Recent criticism towards the use of global-mean regression before analysis has made the degree of anti-correlation hotly debated (Chai et al., 2012; Chen et al., 2012; Fox et al., 2009), but the basic principle of sub-networks showing cohesive changes or decreases in activation still stands. Some have raised the possibility of resting activity amplitudes being more closely related to structural aspects of the vascular system, rather than brain grey matter (Vigneau-Roy et al., 2014), and this may bias results in certain regions, particularly those implicated in the original default mode network. Recent work suggests that the global signal is coupled to local neural dynamics, which would seem to suggest a neural signature. Regardless, even those regions that have a higher flux in blood oxygen level intrinsically may show changes in dynamics that reflect functional neural driven processes, and their frequent usage may have guided their adjacency to the vascular system. The method in this study locates this mean activity in the

principal state, allowing the isolated inspection of additional activity.

Numerous studies have found a number of candidate networks which appear to be synchronized during rest. Some aspects of the current study, namely spectral learning of parameters, have some methodological aspects in common with data reduction methods such as PCA or ICA. The vectors found in those studies are in principle similar to those found here, though they are most often across all spatial voxels rather than the anatomically segregated averages used here, and are without the overarching goal of learning temporal parameters for a more complex dynamic time-shifted model. While using a different granularity of the regional topology will no doubt influence some of the results, the brain is largely organized hierarchically and increasing spatial dimensionality should maintain the general results found here (Power et al., 2012). Few studies have begun to venture into more dynamical models of brain activity, and the precise nature of how resting state activity is related to functional networks is still debated.

While early studies focused on the default mode network, additional work has shown that many other subnetworks are identified during spontaneous activity. These resting state networks, which have been observed across various conditions, including anesthetized primates (Vincent et al., 2007), have been placed under various classes, including but not limited to (Fox & Raichle, 2007) the aforementioned DMN, visual, auditory, dorsal attention and ventral attention/salience networks, motor, and a hippocampal or memory network. These include many regions that would typically have coincident activity during relevant perceptual or cognitive tasks, and there is evidence that there may be re-organization of resting state activity over development (Fair et al., 2007). Perhaps the most important systems in relation to the default mode are the dorsal and ventral attention networks, and if there is interplay between internalized thoughts coordinated through

default mode activity and shifts to external focus, these systems would participate in mediating this change. While these attention networks were originally defined in tasks related to external stimuli, they have also been recognized in resting state studies absent of any demands (Fox, Corbetta, et al., 2006). Both networks are fronto-parietal, with the dorsal being bilateral and proposed to mediate top-down voluntary attention, whereas the ventral system was right lateralized and involved in detecting unattended or unexpected stimuli to shift attention. The principal regions of the dorsal network include the intraparietal sulcus, frontal eye fields, while the ventral network comprises the temporo-parietal junction and ventral frontal cortex, and primate homologues of the ventral network are more debated. Both resting groups have systems that may indicate a right lateralized cluster of fronto-parietal nodes (Figure 4.3A) that coincide with this ventral attention network. These two networks do not seem particularly strongly linked, as represented in Figure 4.3D, and the most common state, involving the default mode, had no particular states strongly predicting it. State 12, which had very strong clusters centered on either caudal or rectus gyrus, was strongly coupled to states 4 and 5, which had either strong fronto-parietal or strong fronto-temporal and subcortical connectivity. State 4 and 5 were directed at state 11, which has strong temporal-parietal connectivity. This may suggest a flow of connectivity in these systems. The visual network is clearly organized in several states among the groups, such as in state 3 where there is a very strong coupling within visual regions. Importantly, state 7 appears to show some ventral visual stream connectivity, and this is the state that is most increased in importance in Dataset 3, where there is the introduction of visual stimuli that would most influence that subsystem.

This intrinsic activity, once viewed as noise, may be a fundamental baseline that is modulated during tasks. Some work has in fact shown that task-evoked ac-

tivity seems to be distinct from rest, but additive in nature, superimposed on some baseline activity (Arfanakis et al., 2000; Fox, Corbetta, et al., 2006). Spontaneous activity can be used to predict individual differences in the degree of learning in a task (Ventura-Campos et al., 2013). Fluctuations in somatomotor cortex activity can also account for the majority of variability in the force of a button press during a task (Fox & Raichle, 2007)., while spontaneous activity in an anterior temporal network is correlated with memory tasks (Gour et al., 2011). As mentioned, Dosenbach et al predict subject age from individually weighted connections from an analysis of brain-wide spontaneous activity (Dosenbach et al., 2010). This all provides evidence towards the view of a basal state of activity that is not stationary, but dynamic and predictive of tasks and behavior, and is itself modified over development. If individual aspects of this spontaneous activity are so variable, then averages across long spans of time cannot be giving a complete picture of the system, but only a reflection of those largest or most common subsystems. Likewise, while the handful of identified brain systems has been a powerful addition to the theoretical toolkit, there has not been a clear mechanism to define how these large spatially distributed systems may account for small individual variations. The idea of overlapping networks that superimpose in time to create observed activity provides a useful template for understanding changes in activity, and our method of untangling complex covariance structure may reflect a deeper relationship intrinsic to regional interactions. This method could provide insight into mechanism and a new approach to understanding the diversity of observed signals in fMRI activity. Some recent evidence from EEG studies (Mehrkanoon et al., 2014) provide additional evidence for this view, where 7 distinct networks were found that were expressed in different mixtures over time.

Two principal kinds of connectivity have been evaluated in brain networks, non-

directional functional connectivity based on correlations, as well as directional effective connectivity typically based on dynamic causal modeling (K. J. Friston et al., 2003) or Granger causality (Roebroeck et al., 2005). Partial correlations have been considered a functional metric with some effective properties, as they more directly tie together relationships between regions (Marrelec et al., 2006; Smith et al., 2011). Functional connectivity is also often interpreted using graph theoretic measures which seek to identify certain organizational principles present in the network, such as path lengths between regions or hubs of activity. These are not dynamic metrics themselves, though they could be evaluated in a subject over time, likewise the set of states found here could be evaluated for distinct network characteristics. Effective connectivity attempts to show some causal effect between regions, though some methods have struggled to reproduce known connections in simulation (Smith et al., 2011). Within attentional networks Granger causality was applied to identify the top down influence of Intraparietal sulcus and frontal eye fields on visual areas during spatial attention (Bressler et al., 2008). Likewise, DCM has been used to show directed influence from bilateral IPS to visual cortex depending on the direction of visual attention, pushing and pulling neural activity in those areas (Vossel et al., 2012). Due to parametric constraints and general difficulty in model construction and interpretability, these approaches typically only look at a small number of regions at once, and while identifying causally directed influence between regions, they assume this influence is stationary over time.

The body of work resulting from static time averaged connectivity has led to considerable insight in brain function and organization, but evidence increasingly points to this being an oversimplification. These experiments tend to span minutes of data, and given the clearly very dynamic nature of neural systems it would be unlikely that there is no change in most metrics used to study the brain over such

a span of time. Resting state activity may have originally been thought of as noise, and now there may be a realization that the dynamics of that activity are also much more than noise. In a recent discussion, Hutchinson et al note a number of issues for this emerging field (Hutchison et al., 2013). The first is an issue that all fMRI studies face, physiological noise and pre-processing artifacts, though these could be of even more concern in studies looking at temporal features. While I haven't addressed this specifically, the reproducibility of our results across three distinct groups, two continuous resting states from different centers with low frequency bandpass filtered data, and a task-positive group with high bandpass filtered data, suggest that the principal findings cannot be due to a specific artifact. Several of their identified issues were specific to most of the methods to date, for instance the use of sliding windows. One strength of our method is that it forms a time-shifted basis that is then regressed on each time point to identify states, without needing a variable window size to create many averages over time. Artifacts may still be present and influence the results, but will not compound over moving averages. Also, while the model may be more complex in theory, it actually has fewer decisions for the investigator and produces a set of readily testable parameters. Design issues include longer scans, and in particular, faster image acquisition, and while our results suggest a set of very consistent basis, some dynamics may be difficult or impossible to detect without faster sub-second scan times.

The literature on dynamic fMRI, resting or otherwise, is still relatively sparse and immature. A number of studies have identified significant changes to connectivity in time, but the complexity and lack of tools for quantifying them made conclusions difficult. Sets of functional connections appear to change between modules over time, often simultaneously and involving DMN or fronto-parietal areas (Zalesky et al., 2014). The modularity of nodes is also dynamic during a learning task, and

that degrees of modularity can be predictive of learning in future sessions (Bassett et al., 2011). Sliding window approaches have shown that there are changes in functional connectivity, with an approach based on principal component analysis (Leonardi et al., 2013) showing some differences in DMN connectivity between controls and subjects with multiple sclerosis. Perhaps the most methodologically related study (Eavani et al., 2013) applied a more conventional HMM approach with sparse basis learning to identify 6 discrete HMM states which are mixtures of a set of covariance basis. This approach limits the basis by assuming small numbers of ROI's change between states, that is they belong to low rank matrices. This helps constrain the difficult learning problem. Additionally a large weakness of traditional HMM may be seen as the result will be identified as switching between some number of discrete states, which is unlikely to represent the true system well unless the number of states is arbitrarily high. These discrete states are good at modeling those systems with competitive inhibition between states, where the observations are either on or off, and this is unlikely to be a valid representation of brain function. Alternately, I might use methods from Linear dynamical systems (Kalman, 1960) that impose Gaussian requirements on noise, but become overly 'smooth'. The approach used in this study is uniquely able to model both inhibition and smooth evolution (Siddiqi et al., 2009).

The degree that resting activity reflects an actual brain at rest is difficult to determine, and it is undoubtedly true that any conscious individual within the scanner is in fact receiving a multitude of stimuli and often reflecting on different mental tasks. The absolute nature of the question has been partially laid to rest by studies that show resting state activity and network formation during sleep, and under anaesthesia. As speculated by Fox et al (Fox & Raichle, 2007) there may be several conceptual layers to resting state activity. The first may be related to unconstrained

behavior or mental thoughts specific to a subject during a task, and will evoke similar activity to respective task related paradigms. A second layer underlies this and possibly all activity and persists across states, conditions, and subjects. It is likely that the variability of states observed in individuals relative to the group may reflect the former, while states consistent across groups reflect the latter, though no doubt still incorporating the common mind-wandering across groups. A reorganization of states, such as those seen here in the task-positive group (Figure 4.4) can also reflect the increase in utilization of a particular subsystem.

An open question is the degree to which functional connectivity is related to anatomical connectivity, both of which are often imaged during single sessions within a subject. Several studies have indicated that functional connectivity is higher in the presence of a strong anatomical white matter linkage (Honey et al., 2010), but comparing structure to the set of basis networks might show that some states have a higher anatomical relationship than others, and function is oscillating around a structural skeleton, and this remains a possible future area to study.

Many studies have shown differences in resting activity between groups having different medical conditions. Not only are differences measurable across age, but there are disturbances in the network in a number of pathological states such as Alzheimer (Greicius et al., 2004) and depression (Greicius et al., 2007; Seeley et al., 2007). Considerable effort has gone into understanding and modeling differences in activity as they relate to these conditions, and it may be the case that specific states or sub-networks dynamics in time are disrupted. Given the evidence that across groups there is a consistent basis that may have variability within subjects, the underlying dynamics may change across different pathology. Only looking at static measures of connectivity may completely miss those aspects that are most disrupted in these conditions. Additionally, the potential to

simulate disruptions through artificial lesions of parameters in the model, or modify regions in pathological models, might bridge understanding in how function that arises through dynamics over time can be either strengthened or attenuated.

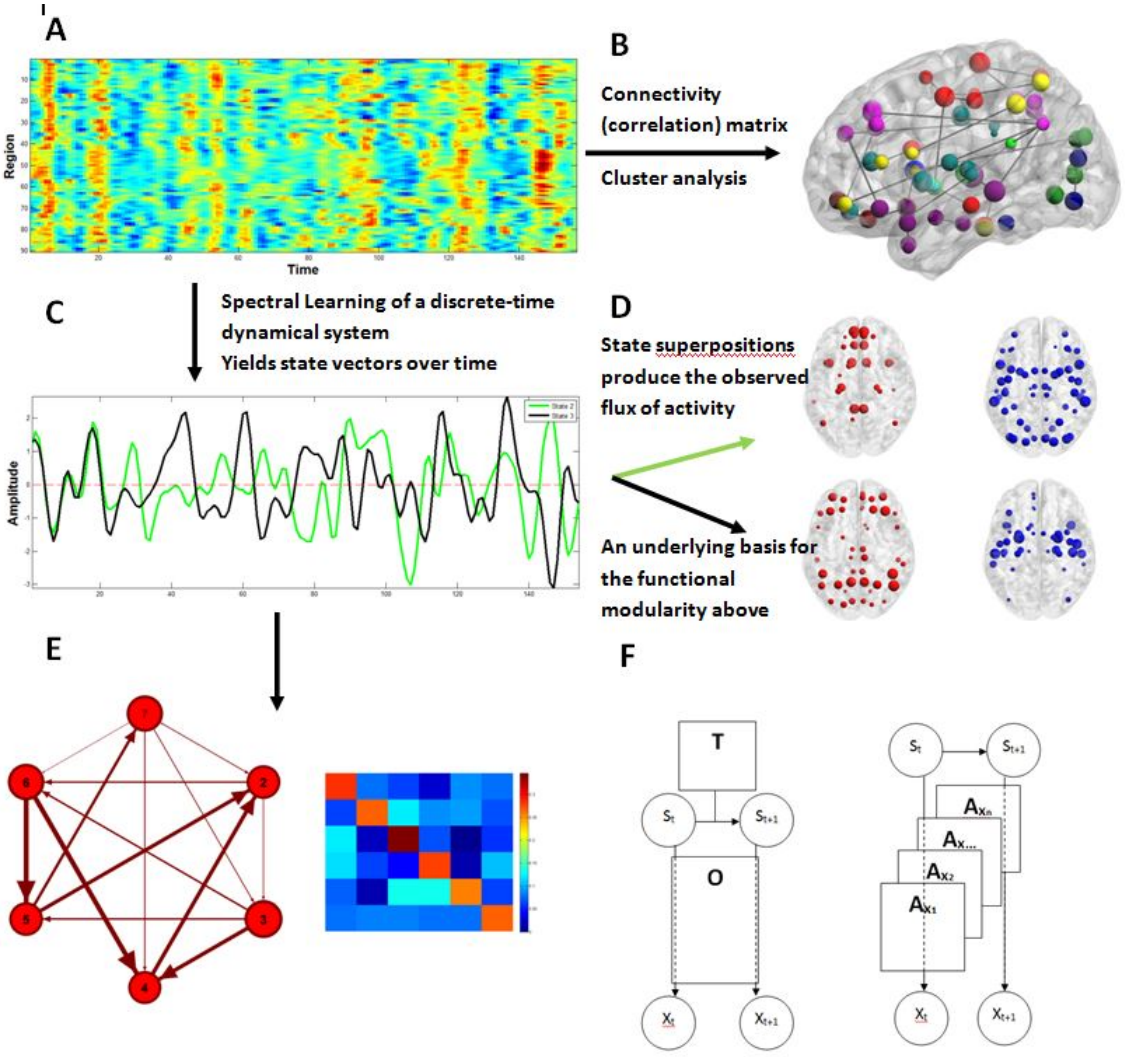


Figure 4.1: A. Timeseries from single subject in Dataset 1. B. An example of a more traditional analysis which generates brain clusters from average functional connectivity. C. This study uses an alternative method to model and characterize the change in internal states of the brain over time. Here two of the most significant states amplitudes are shown over time, corresponding to the time series in A. D Each of these states models the temporal influence of a sub-network of brain regions. Here I show a summation over this structure, and split the result into primarily positive (red) and negative (blue) influence, giving the mean influence of that state, which is weighted by the amplitude in C at a given time to produce observed activity. E I can identify an approximate weighted influence between states, indicating the probability of transitioning between state amplitudes over time. On the left is a graph of the highest weights between states, as seen in the matrix on the right. F This method, unlike Hidden Markov Models, does not explicitly model the transition probabilities T between hidden states S, which then have some emission probability O for viewing observations X, but rather models a set of observable operators A for each observed region, and I can learn state weight vectors purely through linear operations on the set of observables.

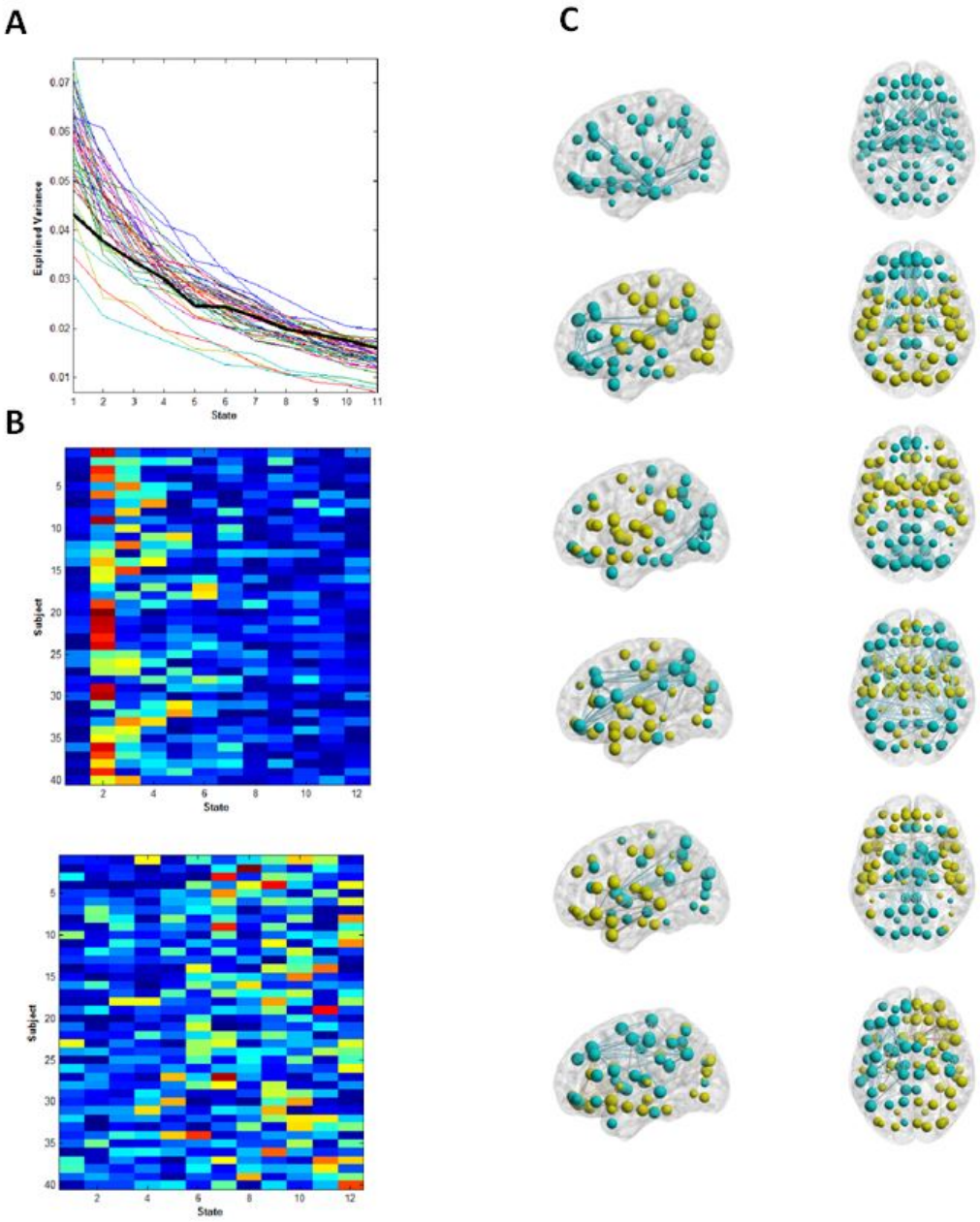

A
B
C
Explained Variance
State
Subject
State
Subject
State

Figure 4.2: A. The systems intrinsic dimensionality can be evaluated through singular values corresponding to each state, and subsequently, an appropriate number of states to use in the model can be determined by investigating their magnitude. Individual subjects have singular values 2-12 plotted, along with the values from a model trained over all subjects in black. The top singular value was not included as it is much larger. B. Individual subject states compared to the group model. The top figure compares each subject states individually to the second state discovered in the group, where below the same is done with state 7 from the group model. State 7 may be a group of regions involved in visual identification and recognition of objects, and its usage could vary considerably between sessions, whereas state 2 is the putative DMN which is frequently identified in resting state activity. C The top six states learned from all subjects in Dataset 1, with sagital and axial views of each. Each map is split into positive (teal) and negative (gold) regions. Edges drawn between regions represent the top 2% of the underlying covariance structure. These distributions represent the major basis networks responsible for most activity during these resting scans.

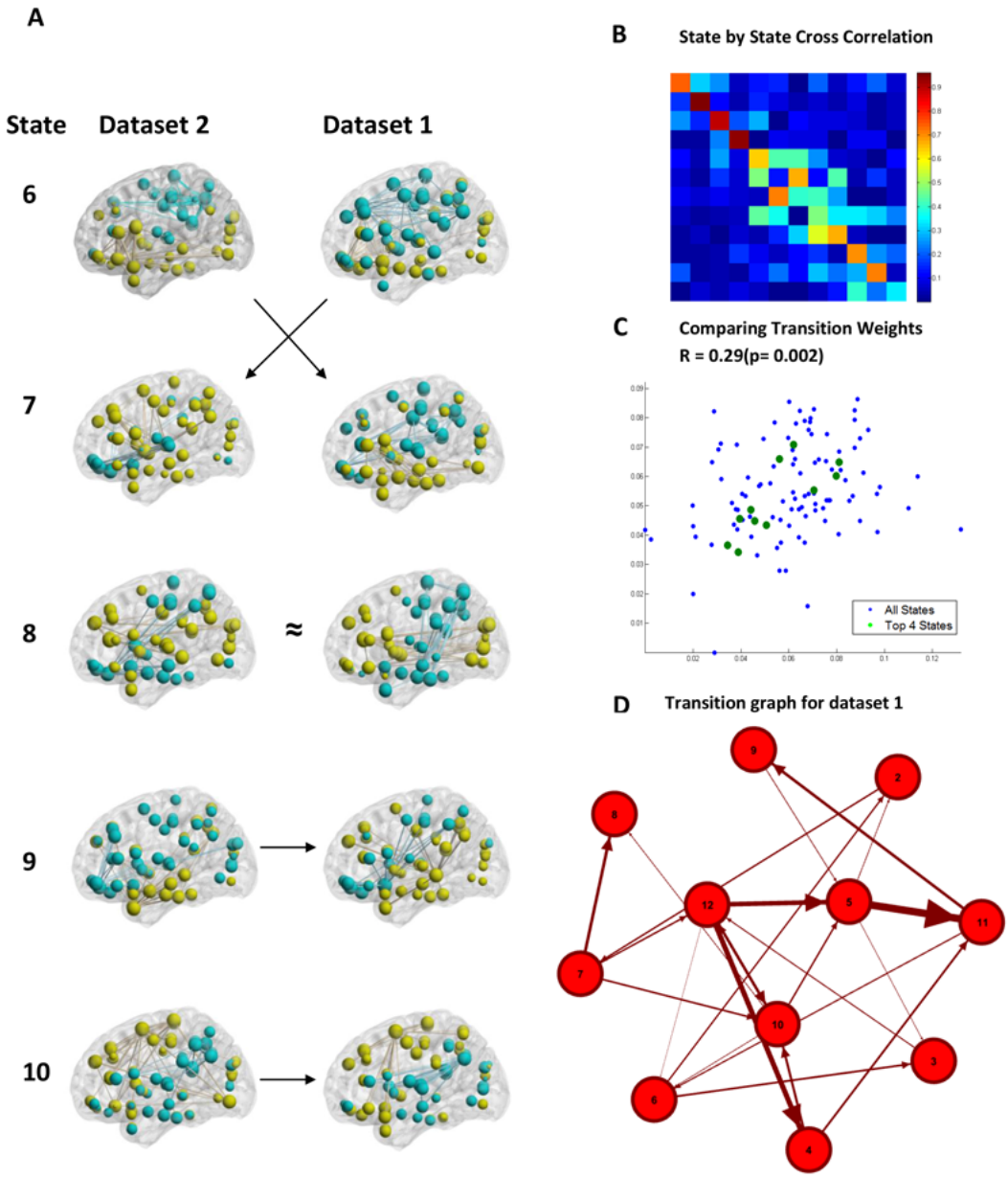

A
State
Dataset 2
Dataset 1
6
7
8
≈
9
10
B
State by State Cross Correlation
C
Comparing Transition Weights
R = 0.29(p= 0.002)
All States
Top 4 States
D
Transition graph for dataset 1
9
2
8
12
5
11
7
10
3
6
4

Figure 4.3: A In an additional public dataset from a different institution models were reproduced with largely similar results. Here sagital views of states 6-10 are plotted as in Figure 4.2, comparing datasets. The distributions of regions are flipped in state 6 and 7, and state 8 seems to be less of a match. B A cross correlation between Dataset 1 and 2 indicates that the first 4 states are very consistent between groups, and the flip in states 6 and 7, and poor correlation in 8 can be seen. C A scatterplot of the similarity of off-diagonal transition weights between Dataset 1 and 2. The dependence between elements from the top states are shown in green and are more tightly correlated. D A graph of transition weight between states in Dataset 1, plotting the top 50% of interactions between states. The distributions of weights is fairly spread out, without very clear organization, though this is also a possible consequence of the method that allows for mixing of states over time. The diagonals of the matrix are not plotted on this graph, which represent a tendency to stay within a state on subsequent time points, this is typically several times higher (as in Figure 4.1E). Edge weight corresponds to transition probability.

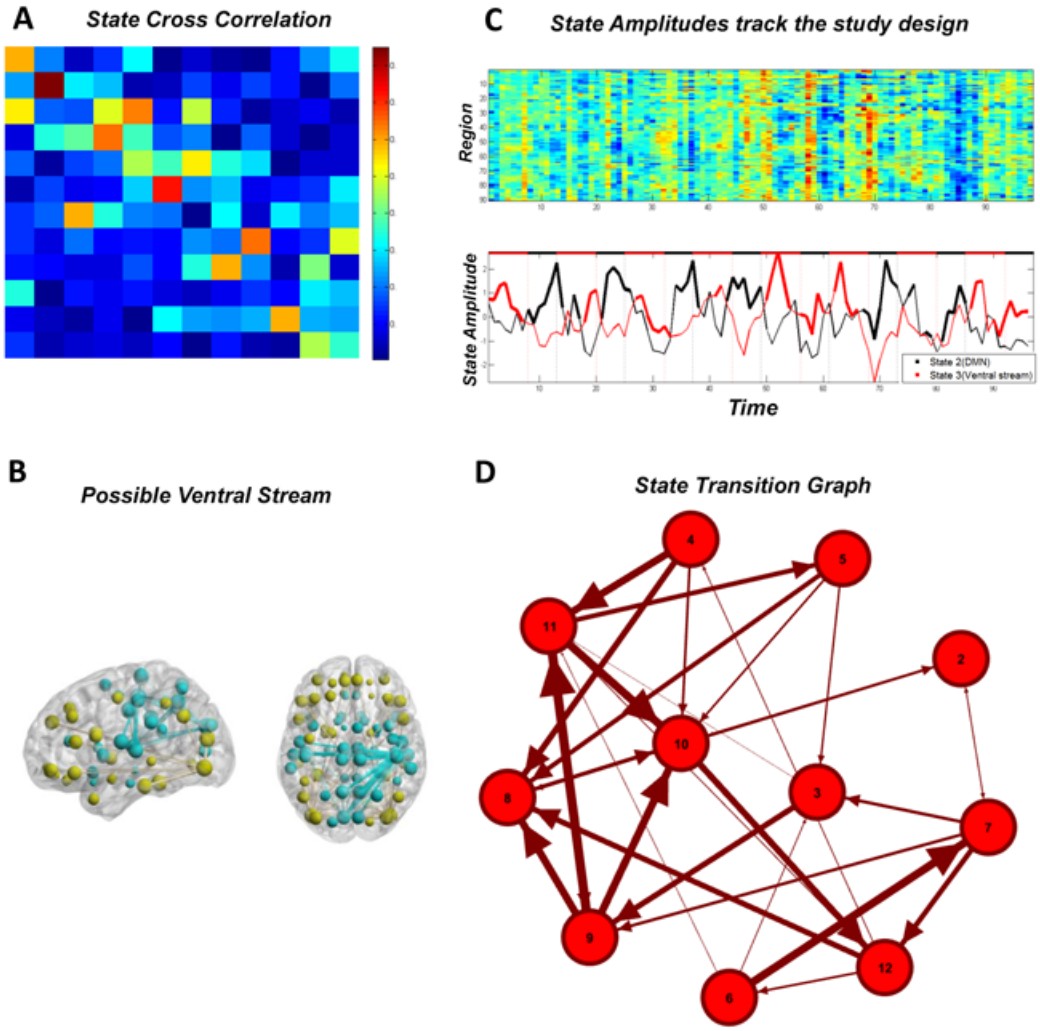


Figure 4.4: A. A cross correlation between Dataset 1 and 3 (compare to Figure 4.3b). While there are still many highly correlated states, they are somewhat less so and do not lie as strongly on the diagonal. This suggests a reordering of the importance of each state on the observed time-series, for instance state 7 from Dataset 1 is most related to state 3 in this data. B. Render of state 3 from Dataset 3, which can be compared to 7 from Figure 4.3a. The gold nodes may indicate the influence of a ventral visual stream, integrating regions from occipital, temporal, and frontal areas. C. This dataset provides a degree of structure compared to the continuous resting state sessions. Above is plotted a single subjects time course over the course of an experiment. Below this timeseries are the state amplitude vectors for state 2 and 3, roughly corresponding to DMN and ventral visual stream maps, and they alternate according to the current state of the experiment which is indicated with alternating red and black bars. D. A transition weight graph between states, where compared to Figure 3d there may be a larger degree of organization into a bistable graph.

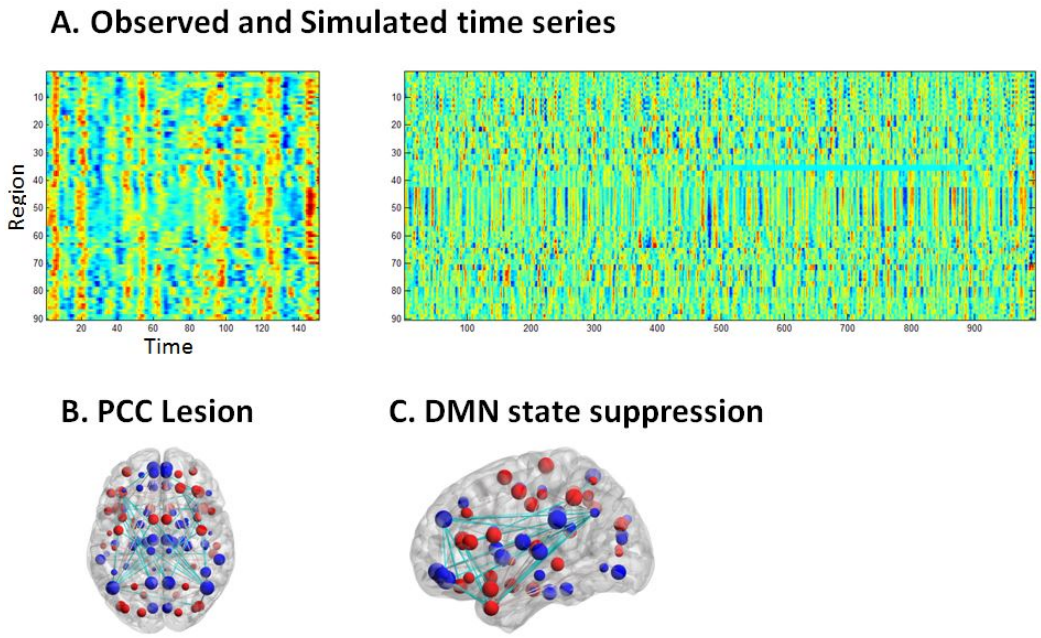


Figure 4.5: A An example time series (left) compared a simulated time course (right). At time 500 an alteration was made to observations, artificially suppressing activity from bilateral posterior cingulate. B Comparing correlation measured from pre and post lesion of posterior cingulate, regions in blue indicate higher correlation prior to lesion, while red indicate areas with lower correlation. The DMN seems to dominate blue areas, though several other states were also negatively affected which incorporated PCC. C Results from the suppression of an entire state over the lesion time course, in this case state 2 corresponding to DMN. Regions in blue show reduction in activity, and include the DMN as well as a number of other regions including inferior frontal and some subcortical areas. Areas with increases in activity are spread out, but the largest changes in correlation are in the temporal pole. The distribution of edges changed through altering this state is qualitatively similar to those found in the next chapter.

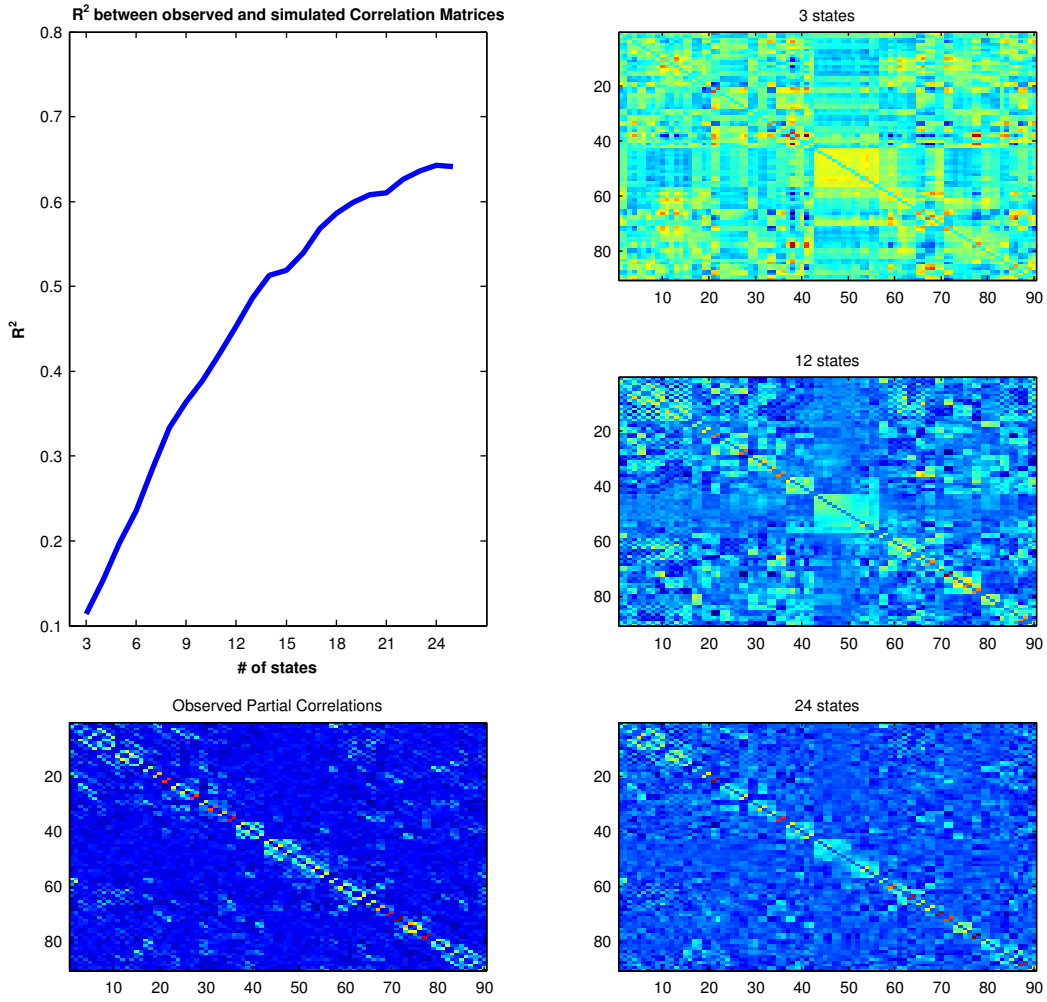


Figure 4.6: Top left: Plot of the effect of the number of states on correlation reconstruction. Bottom left: observed correlation. Right: Several correlation matrices calculated from simulated data derived from either 3, 12, or 24 states. As the number of basis states used increases, the simulated correlations converge to the observed correlation matrix.

	1	2	3	4	5
1+	Temp_Inf_R[2.31]	Hipp_R[2.18]	Temp_Inf_L[1.93]	Hipp_L[1.78]	F_Sup_Orb_R[1.40]
1-	Supra_L[-3.2]	Supra_R[-3.2]	Temp_Pole_Mid_L[-2.2]	F_Inf_Oper_L[-2.0]	Ang_L[-1.5]
2+	F_Med_L[1.91]	F_Med_R[1.83]	F_Sup_Med_L[1.83]	F_Sup_Med_R[1.81]	Ang_L[1.77]
2-	Supra_R[-1.5]	Rol_Oper_R[-1.5]	Supra_L[-1.4]	Post_R[-1.4]	Rol_Oper_L[-1.3]
3+	Cun_L[1.86]	Calc_L[1.85]	Occ_Sup_L[1.83]	Ling_R[1.82]	Calc_R[1.79]
3-	Insula_L[-1.7]	Put_L[-1.7]	Put_R[-1.6]	Insula_R[-1.6]	Pal_R[-1.4]
4+	Amy_L[1.61]	Hesc_L[1.44]	Amy_R[1.44]	Hipp_L[1.43]	Hesc_R[1.37]
4-	Par_Inf_R[-2.5]	Par_Inf_L[-2.3]	Ang_R[-2.2]	F_Mid_R[-1.9]	F_Mid_Orb_R[-1.9]
5+	F_Inf_Orb_L[2.01]	Temp_Mid_L[1.88]	F_Inf_Orb_R[1.85]	Temp_Mid_R[1.56]	Temp_Pole_Mid_R[1.50]
5-	Thal_R[-2.5]	Thal_L[-2.5]	Cau_R[-2.4]	Cau_L[-2.3]	Put_L[-1.8]
6+	F_Mid_Orb_R[2.25]	F_Sup_Orb_R[2.08]	Olf_R[1.91]	Olf_L[1.69]	F_Inf_Tri_R[1.63]
6-	F_Sup_L[-2.1]	PreC_L[-2.0]	Ang_L[-1.9]	S_Mot_L[-1.8]	Post_L[-1.6]
7+	Cau_L[2.00]	Amy_L[1.98]	Hipp_L[1.88]	F_Inf_Tri_L[1.69]	Cau_R[1.63]
7-	Cing_Mid_R[-1.7]	Cing_Post_R[-1.6]	F_Med_R[-1.6]	Cing_Mid_L[-1.6]	Cing_Post_L[-1.5]
8+	Parahip_L[2.11]	Par_Sup_L[1.80]	Par_Sup_R[1.79]	Temp_Inf_L[1.78]	Parahip_R[1.77]
8-	Calc_R[-1.8]	Calc_L[-1.7]	Cun_L[-1.6]	Cing_Ant_R[-1.6]	Ling_L[-1.6]
9+	Temp_Mid_R[2.79]	Temp_Mid_L[1.86]	F_Sup_Med_R[1.67]	Temp_Pole_Mid_R[1.49]	PreCun_R[1.29]
9-	Olf_L[-3.2]	Olf_R[-3.0]	Rect_R[-2.2]	Rect_L[-2.2]	F_Sup_Orb_L[-2.1]
10+	S_Mot_R[2.64]	S_Mot_L[2.48]	F_Sup_R[1.84]	Para_R[1.49]	PreC_R[1.46]
10-	Ang_R[-2.3]	Ang_L[-2.2]	Hesc_L[-2.2]	Hesc_R[-2.1]	Cing_Post_R[-1.7]
11+	Post_R[2.27]	PreC_R[2.16]	Cau_R[2.04]	Post_L[1.96]	Ang_R[1.86]
11-	Insula_L[-2.2]	Cing_Ant_L[-2.1]	F_Mid_L[-2.0]	Temp_Pole_Sup_L[-1.9]	Cing_Mid_L[-1.5]
12+	Cau_R[3.09]	Cau_L[3.05]	F_Mid_Orb_L[2.02]	F_Mid_Orb_R[1.87]	F_Mid_L[1.23]
12-	Pal_R[-2.8]	Rect_L[-2.5]	Rect_R[-2.3]	Pal_L[-2.2]	Put_R[-1.5]

6	7	8	9	10
F_Sup_R[1.39]	F_Mid_R[1.28]	Parahip_R[1.23]	Cing_Mid_R[1.20]	Cing_Mid_L[1.18]
Temp_Pole_Mid_R[-1.5]	Insula_L[-1.4]	Insula_R[-1.4]	F_Inf_Oper_R[-1.1]	Pal_R[-0.9]
Cing_Post_L[1.61]	Rect_L[1.51]	Ang_R[1.51]	Rect_R[1.47]	F_Sup_Orb_L[1.37]
Occ_Sup_R[-1.3]	Post_L[-1.3]	Occ_Sup_L[-1.2]	Occ_Inf_L[-1.2]	PreC_R[-1.2]
Occ_Sup_R[1.76]	Ling_L[1.73]	Occ_Mid_L[1.71]	Cun_R[1.70]	Occ_Mid_R[1.64]
Supra_R[-1.4]	Supra_L[-1.3]	Pal_L[-1.3]	F_Inf_Oper_R[-1.1]	F_Inf_Oper_L[-1.1]
Olf_L[1.35]	Parahip_L[1.29]	Parahip_R[1.29]	Temp_Sup_L[1.22]	Hipp_R[1.20]
F_Mid_Orb_L[-1.6]	Ang_L[-1.6]	F_Inf_Tri_L[-1.5]	F_Inf_Tri_R[-1.5]	F_Inf_Oper_R[-1.5]
F_Inf_Oper_L[1.5]	Temp_Pole_Sup_L[1.37]	F_Inf_Tri_R[1.30]	Hesc_R[1.30]	F_Inf_Tri_L[1.30]
Put_R[-1.7]	Cing_Post_R[-1.6]	Pal_R[-1.5]	Pal_L[-1.5]	PreCun_R[-1.4]
F_Inf_Orb_R[1.60]	F_Inf_Oper_R[1.50]	F_Mid_R[1.50]	Par_Inf_R[1.41]	Supra_R[1.40]
F_Mid_L[-1.5]	Para_L[-1.4]	F_Sup_Med_L[-1.4]	Temp_Mid_L[-1.3]	F_Inf_Tri_L[-1.3]
Hipp_R[1.52]	F_Inf_Orb_L[1.50]	Temp_Inf_L[1.38]	F_Inf_Oper_L[1.36]	Amy_R[1.34]
PreCun_R[-1.4]	Hesc_R[-1.4]	Hesc_L[-1.4]	Cing_Ant_R[-1.4]	Cing_Ant_L[-1.2]
Temp_Inf_R[1.54]	Par_Inf_L[1.49]	PreCun_L[1.41]	PreCun_R[1.29]	Para_L[1.28]
Cun_R[-1.5]	Cing_Ant_L[-1.5]	F_Inf_Orb_R[-1.5]	Pal_R[-1.4]	Ling_R[-1.4]
F_Sup_R[1.24]	Hipp_R[1.15]	Fus_L[1.07]	Ang_R[0.99]	Temp_Inf_R[0.93]
F_Mid_Orb_L[-1.7]	F_Sup_Orb_R[-1.6]	Par_Inf_L[-1.4]	F_Inf_Tri_L[-1.1]	Post_R[-1.0]
Para_L[1.28]	F_Inf_Orb_R[1.28]	PreC_L[1.23]	Temp_Pole_Sup_L[1.11]	F_Sup_L[1.08]
Cing_Post_L[-1.6]	Rol_Oper_L[-1.6]	Rol_Oper_R[-1.4]	Temp_Sup_L[-1.3]	F_Inf_Tri_L[-1.2]
Cau_L[1.85]	Temp_Mid_R[1.35]	Rect_R[1.25]	Rol_Oper_R[1.20]	F_Sup_Med_R[1.14]
Par_Sup_L[-1.4]	PreCun_L[-1.4]	Cing_Mid_R[-1.4]	Insula_R[-1.3]	F_Mid_Orb_L[-1.3]
Para_L[1.17]	Occ_Inf_R[1.16]	F_Mid_R[1.09]	Cing_Ant_R[1.08]	Cing_Ant_L[0.92]
F_Inf_Oper_R[-1.5]	F_Inf_Tri_R[-1.4]	Put_L[-1.4]	F_Inf_Oper_L[-1.3]	Supra_L[-1.1]

Table 4.1: The 10 greatest magnitude positive (+) and negative (-) regions found in each of 12 states in Dataset 1.

Differential networks: A novel view of brain variability

5.1 Abstract

Recently, the importance of brain variability as an independent measure of function has gained considerable attention. In previous chapters I have frequently investigated aspects of functional organization that are largely based on measures like correlation, but here I will directly investigate how this correlation structure appears to vary or differentiate over time and between different visual states. I consider networks derived from the state of the brain when viewing both images of face and non face objects (both referred to ask task-positive) relative to the task-negative periods when no stimulus is present, as well as a cohort of drug users viewing drug related paraphernalia. I contrast this with subjects undergoing continuous rest, as well as a small group of primates under anaesthesia. I hypothesize that a differential between these states may expose sources of dynamic activity that would otherwise go unseen. This derived network is consistent across subjects and groups, both increases in strength and is predictive of age, and discriminates between faces and objects. Additionally, in adult cocaine users there is increased variability located precisely in regions implicated in disruptions to dopamine signaling, including frontal and striatal regions. This network largely matches a theorized distribution of brain regions involved in the representation of semantic knowledge anchored in a hub located in the anterior temporal lobes. It also largely overlaps dopamanergic pathways that are often described as being involved in signal-to-

noise detection. I suggest that this approach to investigating brain networks may identify a novel source of organization above and beyond the more traditional functional connectivity, which reinforces those studies that have found similar results looking at variation within regions.

5.2 Introduction

As discussed in previous chapters, many studies have discovered the reliable organization in fluctuations of the BOLD signal during rest (Raichle, 2009), where correlations between distinct brain region time series appear to have a high level of intrinsic organization. In the previous chapter a more comprehensive picture of this organization was made, where functional connectivity is not static but shifts rapidly between numerous basis networks. Having established this I'll revisit network analysis from an alternate perspective that attempts to more directly measure variability in these time series. If the previous chapter was something of a top-down approach to modeling variability, this will introduce a simpler method of bottom-up analysis that seeks to evaluate the degree or magnitude of variability between brain regions. While the technical details are simpler, this approach produces a unique picture of brain organization that may be distinct from much of the literature.

Recently the analysis of fMRI variability has become a topic of interest itself, and this signal may contain information largely independent of the traditionally used mean. This originated from the acknowledgement of the importance of looking at individual differences in behavior, which typically vary considerably across subjects, age, and so on. With this discovery it was natural to begin looking at how BOLD signals themselves varied with behavior. Many studies have looked at inter-subject variability in fMRI, but recently intra-subject variability has come into

focus. This could be seen as directly using variability in lieu of the mean signal as a statistical measure. This second moment of the data can be independent of the mean, and two subjects with identical mean behavior may have considerably different variability over time. Likewise, two regions of the brain that have similar mean activity may have considerably different variability. Most studies consider this variability uninformative noise and attempt to average it away. Some recent work suggests that this more granular information may be important and it should also be assessed as an additional dimension, and this addition may better reflect the complex dynamics evident across the brain.

Garret et al (Garrett et al., 2010) looked at the standard deviation of time-series and found they were actually more predictive of age than the mean, which was the first work to directly make this comparison as a within-subject measure. They went on to show that this variability is also a powerful measure to identify relationships in differences in task performance such as reaction time. Additional studies have found that increased variability in nucleus accumbens activity correlated with older subjects who made suboptimal choices in a financial risk taking task (Samanez-Larkin et al., 2010). Increases in variability in parieto-occipital connectivity was identified in blind subjects (Leo et al., 2012), which has been suggested to be a hub in the absence of visual input.

Results such as these led to an editorial (Mohr & Nagel, 2010) remarking on the possible utility in using variability rather than means. They briefly review the use of standard deviation, as in Garret et al, but identify that this measure may be poor when the baseline itself is changing, which is frequently the case. They suggest a related measure, the mean squared successive difference (MSSD). This takes the square of differences between adjacent time points, and so can more naturally take into account shifts in the underlying signal, and was used in Samanez-larken

et al.

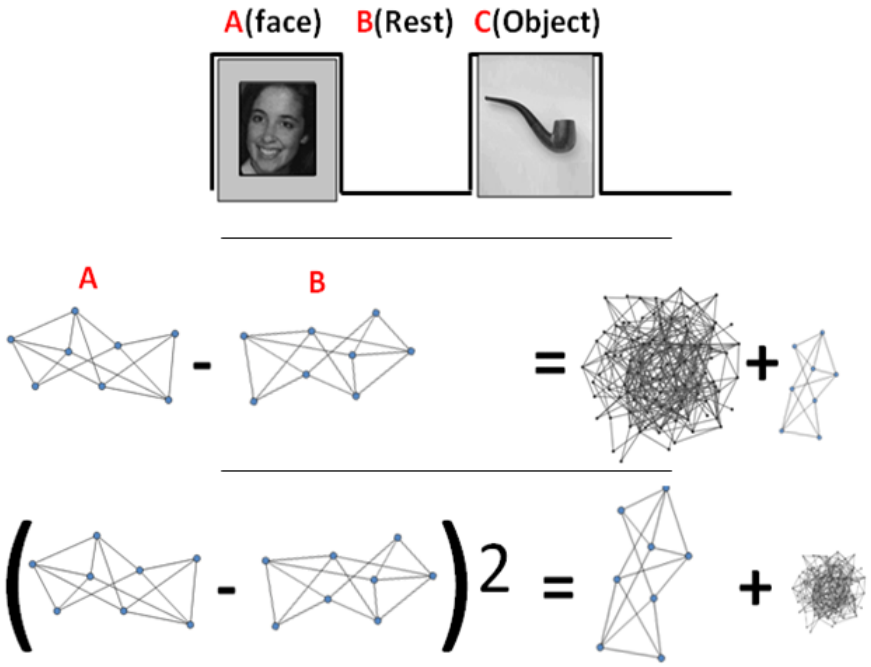


Figure 5.1: Above: the experimental design for task-condition. Over time, different blocks of stimuli are presented to a subject, such as faces (A), intermediate resting periods (B), or non-face objects (C). Correlation matrices, or functional networks, are derived from time points gathered from each kind of condition. Middle: Taking the difference between any two networks would hypothetically yield a result consisting of noise and some remaining signal identifying the difference in conditions. Below: Surprisingly, the square of this difference appears to reverse the functional organization, yielding a consistent pattern of connectivity in areas otherwise seen as noise.

In this work I will investigate sources of variability in fMRI, but rather than look simply at variation about the mean, I will at variability look from a systems or network level. I've discussed in detail the interest in whole-brain organization, and so I'll use this relatively straightforward measure, MSSD, to try to localize what interactions among functional connections have the most variability. As brain networks defined through inter-regional correlations have been demonstrated to be a useful way to investigate fMRI activity, similar to the underlying mean signals, I hypothesize that variation in these networks will be useful and potentially quite distinct

from the original networks, much like variation about the mean signal is. Additionally, this approach may identify putative 'hubs' of variability, possible regional centers coordinating dynamic organization.

MSSD, or what I'll call differentials of functional networks, can be applied easily and rapidly to any time series data. Given the ease of application I will look at a number of cohorts to both establish the consistency of the technique and identify changes between groups. I'll outline results from continuous resting state studies, which could be considered differential networks with respect to time, as well as task-based studies, which could be differential networks with respect to input. Analysis of continuous rest states will allow a baseline of variation to be established, while looking at functional tasks may provide a way to investigate changes in the degree of variability of connectivity. In the task based analysis I'll investigate two groups with very similar designs, one cohort having typical children and adults viewing face and non-face objects, and the other a group of subjects from a study of drug users viewing drug-related paraphernalia. Additionally I will briefly investigate the organization of variability in a small group of anesthetized primates.

This will introduce a novel approach to measuring changes in network organization in the brain, and may also lend itself to some interesting theoretical views of functional organization. I have already identified that the brains functional connectivity is not static during rest, and so I expect to see considerable variation. Likewise, looking at the difference in various task states, such as viewing faces or objects, may reveal how these sources of variability change in different contexts. The distribution of this variation appears to be largely orthogonal to traditional mean activity based measures, which confirms, at the network level, some findings from local activity. This differential network also largely corresponds to a distribution of semantic processing oriented around an anterior-temporal lobe hub,

which has been debated in that literature (Patterson et al., 2007), as well as also largely overlapping with many dopamanergic systems.

5.3 Methods

Participants and data acquisition

The continuous resting state group is the same used in the previous chapter, Dataset 1. The task-positive group passively viewing stimuli is the same as Dataset 2. An additional dataset is used which uses a block design, similar to Dataset 2. This data was supplied from the lab of Dr. Kathleen Brady at MUSC, and it consists of 37 adult cocaine users. In this case, rather than face and non-face objects, these subjects viewed drug paraphernalia and non-drug related objects, and all time-series were preprocessed as in Dataset 2.

For primate data, four female Rhesus monkeys (Macaca mulatta) 6 to 12 years old were habituated to the MRI scanning environment using positive reinforcement. In fMRI scanning (pulse sequence: GRE-EPI; TE=33 ms, TR=2930 ms, flip angle=77°, 64x64 matrix, FOV=10.2cmx10.2cm, interleaved acquisition of 40 coronal contiguous 1.9-mm slices), iron oxide contrast agent was used and heads were stabilized using thermoplastic masks molded individually for each NHP. Anatomical regions (bilateral V1, V2, amygdala, hippocampus, TE and TEO) were defined based on Frey's MRI atlas. Data were bandpass filtered from 0.01 to 0.1 hz prior to analysis.

Additional Motion Correction

Due to the effect of motion on time-series analysis, even when doing traditional motion correction and subject elimination [power], additional evaluation of volume-by-volume motion within subjects was performed, eliminating volumes where changes in head position pass a given threshold. Changes in six parameters of head motion, translation and rotation along each axis, are used to create an index for framewise displacement (FD) in a subject. $FDi = |\triangle dxi| + |\triangle dyi| + |\triangle dzi| + |\triangle i| + |\triangle i| + |\triangle i|$, where $\triangle dxi = dxi_{t+1}$ - dxi_t, and the alpha, beta, gamma parameters are estimated by converting rotation degrees into mm via displacement on a spherical surface with radius approximately the size of the subjects brain. Those frames with FD >= 0.5 mm were rejected from analysis. To prevent biases in estimates, each subject used the same number of images per condition equal to the minimum used across conditions.

Partial Correlations and Estimates of the Sample Covariance

Partial correlations were used to measure the co-linearity between two nodes while controlling for all other nodes. To address the common problem in estimating sample covariance known as 'small n, large p', where a system has more variables than observations leading to an ill-conditioned, singular, covariance estimate we use a shrinkage approach utilizing a convex combination (Ledoit & Wolf, 2003) of the sample covariance with a target diagonal matrix with an empirical shrinkage factor. This constrained structure results in more accurate estimates and is always well conditioned, and thus suitable for matrix inversion and calculation of partial correlations.

Differential networks

Partial correlations provide a first order measure of co-activation of brain regions. In order to investigate variability of these matrices I'll define a differential network with respect to time as the mean squared successive difference of respective matrices $D_t = (D_{t_n} - D_{t_{t+1}})^2$ and $D = \frac{1}{T}\sum_{t=1}^{T} D_t$ Individual matrices are averaged from the local time points, where I use a window as large as minimal block sizes (18 points) in task-related designs to keep samples consistent, and use matrices spaced to avoid overlap. So a continuous resting state block with 180 time points would generate 10 differential networks, that I average for a final subject estimate. When looking at functional networks, such as those derived from face or object viewing, rather than calculating a difference over time I use matrices generated from respective blocks, so a differential network derived from face and object blocks might be $D_{F-O} = (D_F - D_O)^2$

Support Vector Machines

Support vector regression is an extension to SVM that can be trained on real-valued outcomes such as age, as in dosenbach et al, where an envelope of width epsilon is used to evaluate sample points along a regression line. To better interpret feature weights we use a linear classifier, which produced similar results to non-linear kernels across our dataset and produces weights which are more easily interpreted. All SVM and SVR calculations were performed with the Spider Matlab Machine Learning Toolbox (http://www.kyb.tuebingen.mpg.de/de/bs/people/spider). Leave-one-out cross validation was used to ensure generality of the predictive models, where one unlabeled test subject in turn is predicted from a model trained on remaining labeled data.

Hierarchical Clustering

Group averaged matrices were submitted to a hierarchical clustering analysis (Matlab Stats toolbox) using ward's minimum variance method, which minimizes the total within-cluster variance in an agglomerative clustering assignment. This is a deterministic method that produces a tree structure relating groups of nodes to each other according to their shared distances. This tree can be cut to produce as many branches, or clusters, as desired.

5.4 Results

In previous chapters I've described general organizational properties of the brain in detail, such as clustering behavior or latent states. Most of the data analyzed here was also used for either resting state or task-based analysis in those chapters, and so results can be compared easily between different analysis. Broadly speaking, prior results both from my work and the field in general conclude that brain organization typically using correlations is organized into a number of consistent sub-networks that change only slightly during perceptual or cognitive demands. In contrast, the results I find using analysis of variation identify a newly distributed network that appears distinct from those found previously.

In Figure 5.2 I plot a comparison of different networks generated from these measures. The distribution of variation in both resting states (Figure 5.2a) and task-positive (Figure 5.2b, face-rest) is quite similar. There is a higher magnitude of variation in some occipital and temporal regions in the functional condition which is likely related to the impact of visual processing. In both cases, there are strong clusters of activity incorporating anterior temporal lobe with other cortical and subcortical regions, particularly from inferior frontal cortex. This is largely shared be-

tween both conditions.

This distribution of variation can be compared to a network more related to mean-activity changes in 5.2c, where the face-rest condition generates a map of mean level magnitude differences between regions. The underlying matrix used to generate this used differences of signals between regions rather than correlations. This is not at all a common approach to investigating brain organization, but would be related to maps of significantly different regions or voxels in traditional studies. In this case the map is clearly dominated by signal from the inferior occipital, as well as some other visually related regions (such as the fusiform). This is a not-surprising result as these are the commonly identified areas that are face or stimulus specific. Interestingly, this differential network is distinct and nearly orthogonal to that derived from differences in correlation.

I also investigated a group of 4 anesthetized primates in a design mimicking a continuous resting state with no stimulus. A full comparative study would be too complex for this work, but I wanted to investigate general principles of organization in an animal model, as well as possibly investigating how a lack of conscious activity may be reflected in brain organization. Looking at average brain networks (Figure 5.3a) derived from mean correlations there is an organization of clusters that is not unlike that found in awake humans, with distributions of related nodes largely organized bilaterally and grouped functionally in localized anatomical space. In this case there is some higher degree of lateralization compared to humans with several modules being left or right aligned. Looking at the map of variation (Figure 5.3b) there seems to be much less in common with human networks. There is a set of mainly limbic nodes that seems strongly connected, but variability is generally quite diffuse over the rest of the brain. There seems to be two large, widely distributed clusters, which do seem to have some consistent structure such as bi-

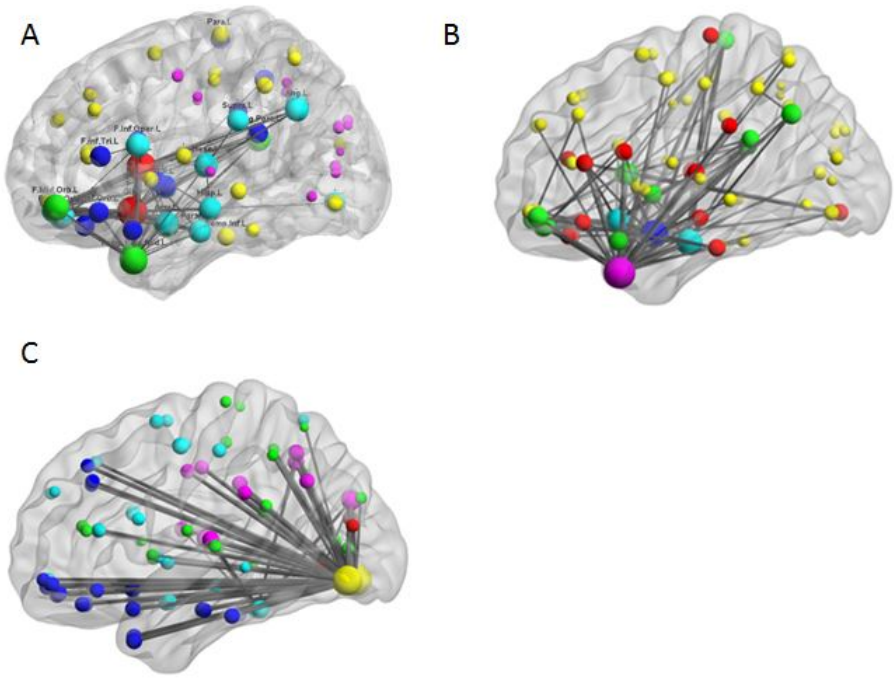


Figure 5.2: A. Using continuous resting data, sliding windows are used to generate differential networks in time. An average of these, over subjects, yields a distribution of nodes centered around inferior frontal and anterior temporal regions, and extending into parietal areas. In B, a similar analysis is done but taking a difference between task-positive and negative times. The network is quite similar to that in A, but possibly more diffuse in nature with additional connectivity notably in more visual regions. In C an example generated from measures more similar to mean-activity analysis produces a map completely localized in inferior occipital and fusiform, generated from the same data as in B. All figures only show approximately the top 2 percent of edges, and node colors correspond to labels from a cluster analysis, node sizes are proportional to strength or the sum of all edges.

lateral symmetry. The limbic system was also incorporated in humans, which may be the common element in both cases. It is possible that the anesthesia has had a profound effect on variability, while having a more modest one on mean activation.

Various measures of mean activity, brain networks, and sources of variability have been used to match or predict individual subjects ages, and I have been able to replicate that here using these new measures. In Figure 5.4a I show the average brain organization from mean correlation matrices across subjects when

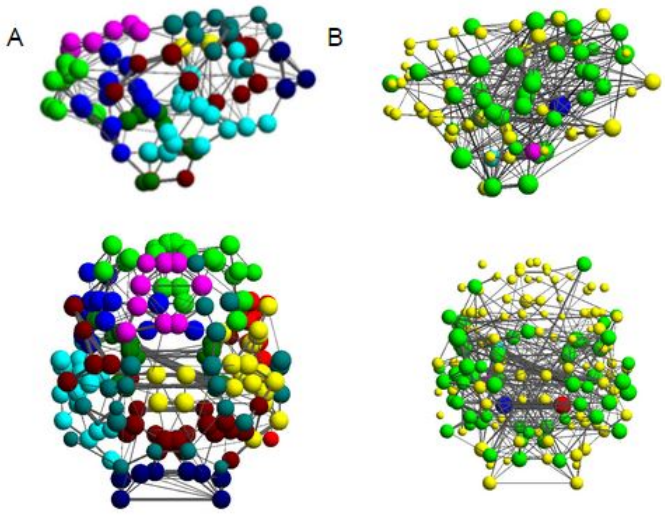


Figure 5.3: Here average data from 4 primates was analyzed as before. A. Graphing the average functional organization from correlation matrices produces a distribution of nodes not unlike those seen in human data. Organization is largely localized anatomically, with considerable bilateral symmetry and functional organization, for instance in visual regions. B. When plotting differential networks, there is some structure in organization of variability. Modest bilateral symmetry seems to discriminate between two major components in yellow and green, with a small subnetwork of more limbic nodes being separated from others. This organization is in stark contrast to the more typical functional organization, and seems quite unrelated to human data.

viewing faces. Using this I can use a support vector regression to accurately predict age, and any similar functional map, such as those generated by task-negative or continuous resting state subjects, also successfully predict age. In this case it appears that younger subjects may have a relatively poor separation from each other, before a developmental curve picks up in teens. In Figure 5.4b average face-rest differential organization is plotted next to a similar support vector regression. This results in nearly as good of a prediction as those generated from the mean correlation matrices. Additionally, a very simple sum over all squared variance results in a crude but significant age prediction. Below in 5.4c is a similar result when looking at a face-object map. Finally in 5.4d I plot a result from summing the squared difference between randomly selected time points, that is a difference from

networks generated from mixes of face, rest, and object viewing. This also results in a significant linear age relationship, suggesting that the principal factor driving the age relationship is a general increase in variability independent of function.

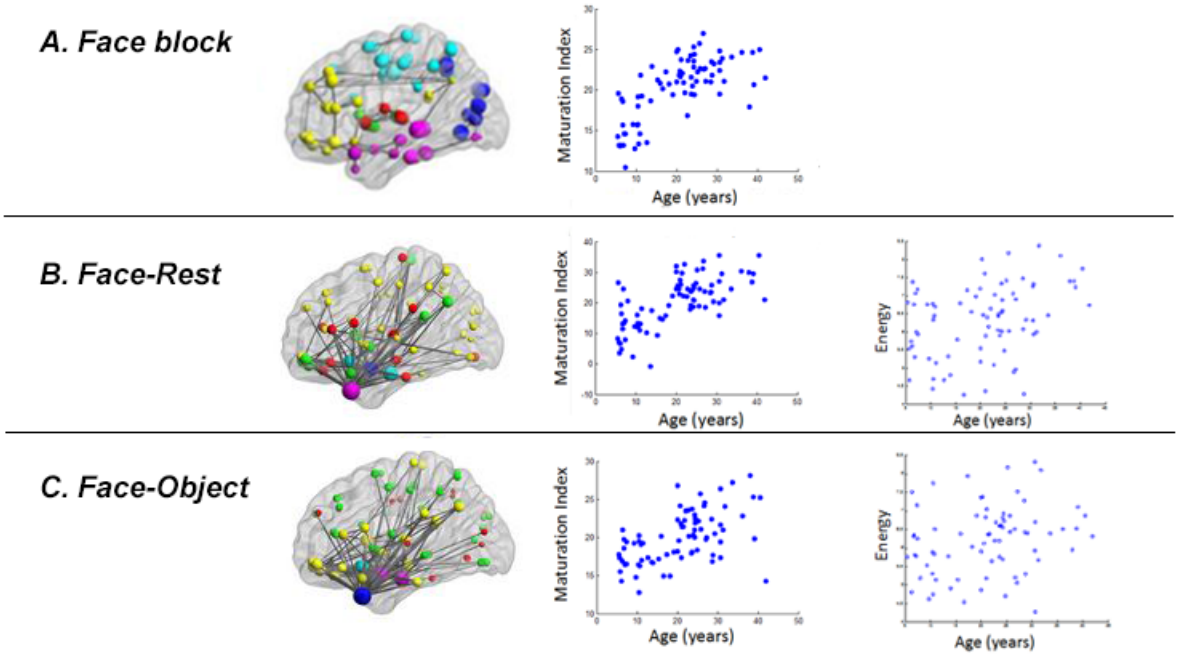


Figure 5.4: A. Correlation matrices generated from face-viewing blocks produce a map of functional organize (left), and can be used to create a maturation index that predicts subject age (right). Predictions are done through a support vector machine analysis, where models are sequentially trained on all subjects but one who is used to test. Similar results can also be obtained from object, task-negative, or continuous resting state. B. Face-Rest networks can also be used to generate such age predictions, with what may be a slightly different growth curve. C. The same analysis but done in a face-object differential network, which also produces a linear maturation index, though it may be more dis-similar than the previous growth curves. Right column in B,C. Differential energy vs subject age, which is simply the sum over all magnitudes of the differential matrix. This also yields a significant age relationship without the need of any prediction or machine learning, something not present in the simple correlation data.(Face-Object R=0.35, Face-Rest R=0.25)

Despite the similar distribution of variance across all conditions, there does seem to be distinctions between groups or stimuli. In Figure 5.5 I show the results of a support vector classification of networks derived from face-rest and object-rest.

Adult accuracy is 87% and across all subjects is 75%. The significant features discriminating conditions do seem to have some considerable bi-lateral distribution, though they lie across the brain rather than mainly within the principal fronto-temporal hub. Using edge weights as a measure of influence, the face differential seems to be more localized, with some notable regions including right fusiform and occipital regions, bilateral posterior cingulate, left middle cingulate and amygdala, and left inferior frontal. The object differential seems less localized.

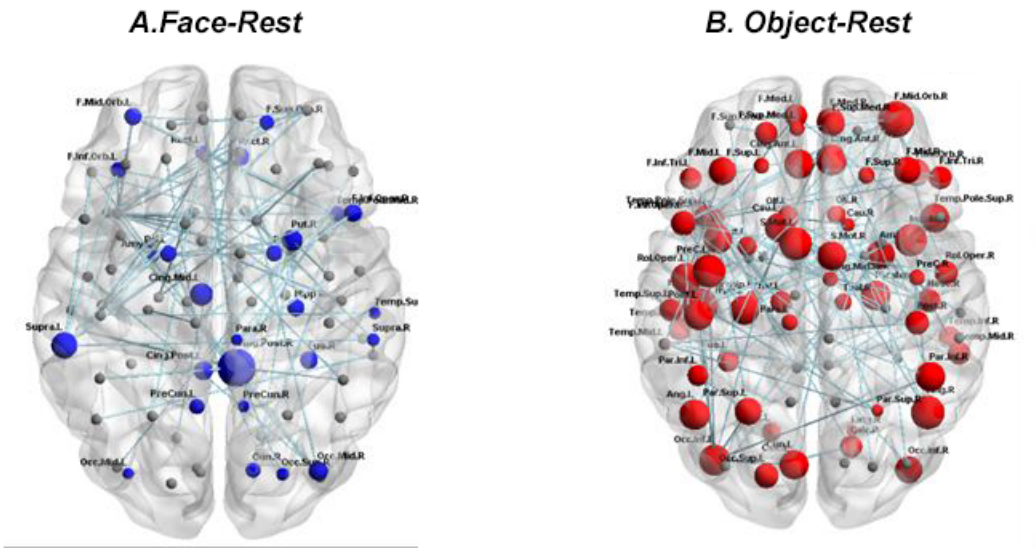


Figure 5.5: Using a support vector classifier, networks derived from face-rest and obj-rest were successfully classified using a leave one out approach. A. Edges that were predictive of face viewing. The edges and nodes most predictive of face viewing. These can be interpreted as regions and interactions displaying greater variability between rest and face viewing. Node size is the sum of edges. B. Edges predictive of object viewing. These seem less concentrated than those discriminating faces.

An additional group of subjects was analyzed with a design very similar to the face localization task, but rather than compare faces and non-face objects the design was looking at drug users viewing drug-related periphenalia. In Figure 5.6a I plot the highest sources of variability within this group, and in contrast to the typical subjects viewing faces and other objects, there is a pronounced difference

in network variance. In Figure 5.6b I directly plot the matrices generating these Figures, comparing a typical group viewing faces to the cocaine users. A number of regions have much higher variance, particularly interactions among amygdala, inferior frontal, caudate, and putamen.

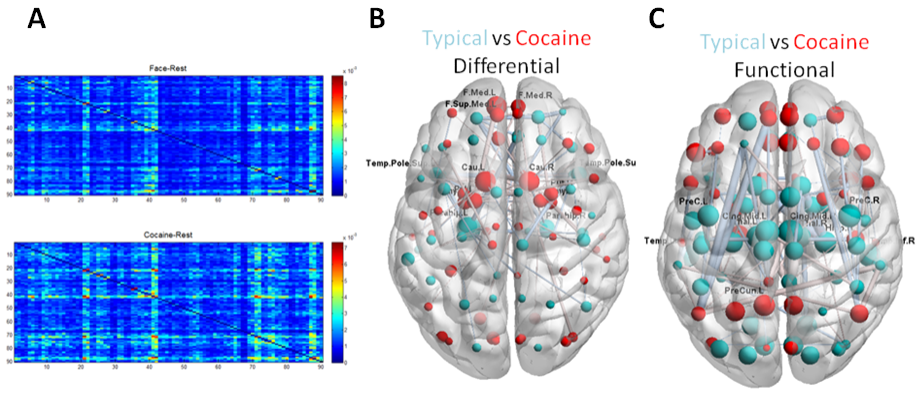


Figure 5.6: A. Matrices generated from group averages of the typical subject's face-rest condition, and the cocaine users drug image-rest condition. While the network organization is largely the same across both groups, some parts of the cocaine group have much higher deviation. B. Brain regions with higher differential network magnitude are indicated in red for cocaine users, and in blue for typical users. There is a strong concentration of increased magnitude around the medial frontal, amygdala, and caudate. C. A comparison figure generated from traditional connectivity maps, showing that while there are large deviations between groups they are organized much differently.

5.5 Discussion

In this study, I attempt to identify and characterize patterns of variability in functional connectivity in BOLD activity. I show that these differential networks are useful measures of brain function, acting as an indicator for subject age, as well as discriminating between classes of visual stimuli. Further, I identify possible evidence for disruption of this differential network in drug-users. This suggests that there is a specific spatial pattern of temporal and task-specific variability found in

functional networks, similar to recent results suggesting that variability of the BOLD signal itself is a useful, and largely independent, metric. It appears that this is an additional kind of functional representation that exists independently or in addition to the more traditional functional organization frequently studied in fMRI.

Interpretations of variability in fMRI

Samanez-Larken et al (Samanez-Larkin et al., 2010) investigated individuals BOLD variability and their degree of risk taking. They hypothesized that older adults might be impaired in risky decisions due to age related declines in dopamine, which lead to more variable signals in the striatum. They suggest that this variability might compromised value estimation. In particular, they looked at the temporal variability in nucleus accumbens which is a region identified in various risk-seeking tasks. Using a mediation analysis they find that temporal variability in NAcc increases with age, and this variability causes differences in risk seeking mistakes. On the other hand, the mean signal in this region does not predict the same mistakes. There is a general theory of dopamine function that suggests dopamine influences the signal-to-noise ratio of information processing, effectively sharpening brain representations and performance. This is one hypothesis for why variability in BOLD is tied to age, as dopamine levels decrease during normal aging.

The core regions that are involved in the differential networks found throughout this work are also largely concentrated around the various dopaminergic pathways. Mesolimbic, mesocortical and nigrostriatal projections tie together limbic, frontal, and striatal regions, which effectively spans the same regions as these differential networks. Additionally the drug-abuser cohort found distinct increases in variability in frontal and striatal areas commonly affected in drug use due to dopamine disruption. This suggests one potential mechanism underlying these differences in variability, where moment to moment changes in dopamine may be reflected in the

underlying temporal variability of BOLD activity.

A more general and theoretical view of what BOLD variability may represent comes from some earlier work. McIntosh et al (A. R. McIntosh et al., 2008) found that multichannel EEG variability increases in adults may reflect a more sophisticated neural system that is able to explore multiple functional states compared to children. These increases in variability they found were also coupled to faster reaction times and better performance, suggesting a role for these fluctuations in optimal behavior. This line of thought has been theorized across work that suggests that variability over trials or time might derive from underlying changes in coherence or spontaneous activity (Laskaris et al., 2003; Nir et al., 2008). That is, regions that have a real functional coupling to each other will exhibit greater variability as their co-activation exhibits itself, and this activity and coordination between neural systems might be better seen in their variability rather than mean activity. Under a number of conditions, theoretical work suggests that neural networks can spontaneously organize between up and down states of activity (Parga & Abbott, 2007), and these naturally balanced systems may be nonetheless inherently variable. Importantly, these analytical systems were found to be sensitive to outside noise, driving modulations into up and down states.

This evidence may provide a systems view of the functional and active brain, where the kinds of organization seen in the dynamics from the last chapter are a result of a complex balancing of neural systems interacting with each other into the multi-state system. These results all point to an even more dynamic brain that thought, and simply using mean activity will miss most of this information. Certain subsystems, such as the anterior temporal lobe identified here, may hold a special role in coordinating this activity. This differential network also overlaps with many of the key dopamanergic regions of the brain, which has long been seen as playing

a role in controlling the signal-to-noise of information encoding. Greater variability seen in some of these regions with age may reflect an increase in the underlying functional connectivity, that is with age these neural systems are more highly integrated with additional brain regions. This causes the fluctuating activity to shift more rapidly in time as it goes in and out of coherence with various other systems. This is related to a proposed mechanism for the changes in BOLD activity and inter-regional correlations found through normal aging (Fair et al., 2009), where it was found that brain organization follows a progression from locally organized anatomically defined clustering to more distally organized functional clustering, for example the appearance of a more strongly correlated default mode network linking frontal, cingulate, and parietal areas. In general the increase in variability may be seen as a consequence of the introduction of additional states that the brain can occupy, and is a reflection of greater computational and behavioral flexibility. The variation of fluctuations between these states is likely happening at a scale much faster than the recorded fMRI activity, but are nonetheless reflected in these observations.

Garret et al have found that the changes in variability across the brain during normal aging are mixed between areas that both increase and decrease in variability over time. This would challenge a naive idea that aging, and its related decline in many tasks, is completely a function of noise as a system 'decays'. Rather this noise is a more complex component that is a reflection of changing brain organization. Those systems that have reduced variability through normal development may reflect reductions in adaptability observed after adolescence. In contrast to previous investigation of variability this chapter has not looked at a regions specific signal, but variability of inter-regional correlations. This allows a more direct comparison of the functional coupling often discussed more theoretically, rather than

an indirect look at how an increase in coupling might be reflected in variability of the underlying signal.

McIntosh et al proposed an inverted U of brain variability across age, that can be matched to a U shape curve of behavioral variability across age. That is, children have a great degree of variability in their behavior as their underlying neural systems have limited coherence between functional systems governing behavior. Through development, these systems better integrate, are able to flexibly move between more states, and behavior becomes more controlled. In old age these neural systems may begin to decouple again, and behavioral variability once again increases. This ties into aspects of dedifferentiation theory (Baltes & Lindenberger, n.d.) where older brains become less distinct during task performance through reductions in selectivity and specificity. Along these lines, Garret et al find lower brain variability in older, poorer cognitive subjects.

It has been proposed that the concept of dynamic range, which refers to the range of possible neural responses, be extended to variability. Perhaps this variability might be seen as an analogue for the range itself, where this dynamic range is not any particular measured signal but the set of neural couplings possible. Ultimately the underlying neural system is composed of the connectivity between a large, and at the level of fMRI, unknowable neural populations, and the information expressed in a neuron or population of neurons is a function of this connectivity. The observed BOLD signal is a summation of instantaneous metabolism tied to neural activity, and it is likely that the variation in this signal is a better marker for connective range than the mean. While this work looks directly at a differential metric of functional connectivity over time, it should be noted that many kinds of analysis, such as that found in the previous chapter, have an implicit modeling of these changes in variation as well. Those results may not directly assess the

spatial distribution of variability, but that distribution is governing computation of a covariance structure in time.

Semantic representations

My earlier investigation into these differential networks was performed in the task-positive condition and then later reproduced in a temporal rather than semantic context in resting states. As the distribution of nodes was established as being distributed around inferior frontal, and anterior temporal lobes, the surprisingly good alignment with a network hypothesized to account for semantic representation in the brain was noticed. Semantic memory or conceptual knowledge corresponds to the knowledge of objects, words, and representations in general (Patterson et al., 2007; Visser et al., 2010; Wong & Gallate, 2012). This is in contrast to episodic memory that places or orders events and objects. Abstract representations of information in the brain have been difficult to study as they are typically thought to be distributed in some manner and presumably integrate numerous modalities and systems, ie vision, language, auditory, and so on. One extreme but plausable idea was that a universal connectivity might be responsible for representations. This is the idea the entire cortex and some subcortical regions may be responsible for representations and so no particular stable subnetwork could be identified as coordinating this knowledge. Lesion studies largely eliminated this as a possibility, but the nature of this representation has still been elusive and is likely at least partially distributed.

Most recent work has been built around the idea that semantic memory relates perception and action, and so should be represented or overlap with those regions (Martin & Chao, 2001). This is necessarily a distributed system, but a key ques-

tion is whether the encoding is done in a purely distributed manner or if they are integrated in some hub region. While the numerous features of a representation are encoded in different areas of the brain, the integrated representation might be encoded in a distinct area. Subjectively, it seems that semantic knowledge as we understand it requires one additional step beyond encoding, the ability to generalize concepts beyond the attributes of that given object. The idea of an object can incorporate many sensory or even abstract properties of the entire class of related objects. This allows for the learned representations to be used in novel contexts or scenarios not previously seen. Having an independent hub that was involved in this coordination of properties seems a testable hypothesis for the embedding of semantic information. Not only should it be possible to identify possible hubs, but disruption to that area should impair semantic judgement independent of the type of input.

Considerable evidence has pointed to the anterior temporal lobe as a strong candidate for this semantic hub. Semantic dementia is a neuro-degenerative condition with progressive deterioration of vocabulary and knowledge about everyday objects. By the time these patients become systematic there is typically bilateral degeneration of the anterior temporal lobes. Similarly Alzheimer's patients may encounter similar semantic issues, though the more distributed degeneration in these patients that may extend into anterior temporal regions probably accounts for the inconsistency in effect compared to Semantic dementia. A number of other focal disruptions to ATL, such as stroke, might impair semantic recall. Additionally, the role of ATL was further solidified in a TMS study where symptoms mimicking semantic dementia were induced through low frequency repeated TMS (Pobric et al., 2007). Interestingly, relatively few fMRI studies identify the ATL as a significant region activated during semantic tasks (Patterson et al., 2007). Typically, these

studies find numerous regions in frontal, posterior temporal, temporo-parietal and parietal areas, which are likely constituting many parts of the putative distributed semantic network. Several reasons for this lack of activation have been given. Principal among them may be that the ATL may have relatively low signal-to-noise due to proximity to sinus cavities, and PET may be more able to detect this activation (Devlin et al., 2000). Magnetoencephalography studies have also shown that in a semantic task, activation from sensory areas as first detected with convergence in the ATL about half a second later (Marinkovic et al., 2003).

The network identified in this work is clearly anchored around the anterior temporal lobes, bridging frontal, temporal, and parietal regions. This at first seemed to be a powerful linkage between some kind of variability present in the task-positive task and a previously unseen semantic network. Further analysis of resting states found a very similar pattern of activity in subjects not involved in any specific task. In either case, there were no specific semantic knowledge tasks being presented, rather more passive viewing tasks. It may be possible that subjects are using this network continuously through conscious processing, with evidence of increased incorporation of visual and sensory areas during the visual task compared to rest. The flux between resting and attentional states has largely been used as an explanation for the emergence or disappearance of subnetworks such as the default mode, but the semantic network may be just as taxed when reflecting on internal cues and demands as external ones. The semantic network may also be constantly utilized for integrating the functional modes seen in the previous chapter, and so these regions account for the most variability in functional connectivity while not actually constituting one of the derived networks. Interestingly, the relative lack of this network emerging in primates may suggest it does have a role in conscious activity, and a comparative analysis in sleeping or anesthetized humans would be

an interesting next step.

There appears to be a highly reproducible network that may coordinate or interact with many other frequently seen functional clusters, and this network may have been missed in most fMRI studies because they mainly look at mean measures of activity rather than the variance of that activity. This network may also be related to some theoretical models of brain function, namely a free energy model proposed by Karl Friston (K. Friston et al., 2006). Variation could easily be viewed as an 'energy', where the mean difference is a velocity away from the information-theoretic position of the brain at a given moment. The general increase in this energy with age may reflect the expanding degrees of freedom in the number of states the brain can explore.

Conclusion

In this work I report the traditional functional organization of the brain as observed through inter-regional correlations in chapter 3, as well as a brief investigation into possible developmental disruptions in Autism. This establishes, and largely reproduces, some basic notions of how the brain is segmented spatially through analysis of its temporal correlations. This also reproduces some previous work that was able to use this intrinsic organization as a marker for a subjects age or brain maturation.

After a review of relevant literature in chapter 2, I go on to demonstrate that the dynamics of the brain during rest are not stationary as typically assumed, but vary as different sub-networks co-activate. Nearly all work to date has assumed that correlations were stationary in time, which I show may miss a considerable and important aspect of dynamic activity. To produce this more robust picture of the functional brain I apply a new method from recent machine learning research that estimates latent aspects behind these dynamics using spectral learning theory. When comparing this model to one from subjects involved in a visual task, I find that a ventral visual sub-network becomes more pronounced, and the dynamics between regions changes in a subtle manner that likely reflects the functional reorganization under task demands.

I also introduce a new measure that finds a differential network that may expose the sources of dynamic activity driving the dynamics above. This network

appears to be largely absent in anesthetized primates, and is disrupted in a cohort of cocaine users viewing drug related paraphernalia.

These results suggest a dynamic brain containing interactive systems that appear to exist in a multi-stable steady state. They superimpose to create the observed BOLD signals and re-arrange during visual tasks. The creation of a model of the dynamics also introduces an opportunity for simulation studies, where perturbations of the system may create translational links to atypical groups. Finally, there may be an independent level of organization that is coordinating this dynamic activity, which is identified through an analysis of the systems variability. These results further our understanding of how the brain is organized during rest, and how that structure may reorganize when presented with a stimulus.

Cortical Activity as a network at the edge of instability

This work has reinforced the idea that the spontaneous activity in the brain during rest is structured into a number of particular spatiotemporal patterns. These patterns appear to emerge despite the enormously complex and noisy system driving this activity. This suggests that this complexity is constrained through some sort of equilibrium state. By better characterizing the nature of dynamics about this equilibrium, and identifying a putative network of variation possibly coordinating these dynamics, this work may help to understand the nature of spontaneous activity and its relationship to evoked activity during tasks.

This equilibrium state appears to be organized around the anatomical framework of the brain, as determined through diffusion tensor imaging (Honey et al., 2010). These studies have shown many areas with high functional connectivity also tend to have a strong anatomical connectivity. These include many of the re-

gions in networks such as the default mode, for example the posterior cingulate cortex and precuneus have high connection density, suggesting they are a structural core or hub. While anatomical connections are typically predictive of function, the reverse is not always true. Additionally, these functional connections are typically more variable than the anatomical linkages, even in studies of within-subject retests for reliability (Koch et al., 2002).

If there are functional networks that exist without any stimulation, and are largely independent of each other without being driven mainly through anatomy, how do these networks persist? A purely cognitive explanation, such as modes of introspection, may fail to completely explain them as at least some functional networks exist during anaesthesia and sleep. Deco et al propose that it may be appropriate to consider these resting dynamics as being a picture of the brain in constant exploration, shifting through multiple states to generate predictions of likely network configurations that are optimal for whatever inputs may be coming next (Deco et al., 2011).

This may be a useful hypothesis, but does not simplify the underlying problem - how can we understand and model the collective dynamics in such a complex system as the brain? They discuss several theoretical and computational models that attempt to recreate some of the connectivity seen in static networks; the dynamics discovered in this thesis may provide an important step in bridging our understanding to these computational models. These models are mainly concerned with trying to simulate networks from static anatomical connectivity. This is in contrast to my own approach; seeking to better characterize the modes of dynamic activity through observation. The two approaches can be informed by each other, and eventually begin to constrain the space of outcomes of each in turn.

Briefly, in their review, Deco investigated three models in detail. They applied

various kinds of connectivity based on anatomical connections, with simulated activity derived from these connections driving responses, and they used these to investigate how fast synchronized activity - typically in the gamma range - might drive the kind of slow oscillations seen in resting activity. Using these models a number of system effects similar to observed activity can be seen, such as the emergence of several coordinated subsystems largely spanning frontal and parietal regions, and the anti-correlated nature of these systems. One important aspect of the simulated systems was that the time-window used in capturing data may influence the observed functional networks, as nodes often form distinct networks for relatively small periods of time. This of course was a major factor in looking at time-varying dynamics, identifying how these networks evolve, but it also suggests that the temporal resolution of the scanner may be a factor.

They also warn against the notion of a discrete number of functional networks, suggesting that anatomy provides a deterministic structure as a scaffold for certain functional modes, but that the precise configuration is a function of intrinsic activity (or exploration) and environmental demands. If there is a multi-stable set of brain networks, constrained but not limited to the underlying anatomical connectivity, noise may provide the 'push' or kinetic energy needed for the system to visit network configurations, without being a 'blank slate' system waiting for input. Additionally, through visiting these configurations the system is able to shape and predict future inputs. This provides important links to a large body of theoretical and computational work. In a follow up study (Deco & Jirsa, 2012) they further identify that the best fit to BOLD data is obtained when the underlying network is critical, that is when the brain is operating at the brink of a bifurcation separating a possible stable equilibrium low activity state from the multi-stable state where many 'attractors' cause high activity in different brain areas. I've also found evi-

dence supporting this view, as the superposition of states generally does not stabilize into any particular equilibrium, rather the brain appears to be in a constant flux between modes. These intermediate states may represent the latent 'ghost' attractors proposed by Deco et al.

Bayesian brains and free energy

A number of related theories with considerable experimental evidence supporting them have begun to explain perception and action as being driven through a process of unconscious probabilistic inference (Knill & Pouget, 2004; Pouget et al., 2003). This idea is in some respect very old (Helmholtz, 1910), and might be seen as the linkage from a neuronal information theoretic world to external physical systems, clearly stated by Mach who claimed that 'the foundations of science as a whole and of physics in particular, await their next greatest elucidations from the side of biology, and especially from the analysis of the sensations'. (Ernst Mach, 1897)

Aided through modern statistical and machine learning techniques, researchers have found that humans and animals frequently behave as optimal Bayesian observers. This proposes that the information about the world, as gained through the senses, is represented by a conditional probability over a set of unknown variables (the posterior). A Bayes optimal system maintains a representation of the parameters being computed with their probabilities, allowing a system to integrate information over time from different sensory modalities, and this representation is maintained through stages of processing. In Bayesian networks this concept is known as belief propagation.

The basic premise of Bayesian theories of cortical processing is that the brain

encodes with something like probability density functions, rather than through a deterministic and discrete mode of operation. While a deterministic and discrete picture of the brain is more accommodating to our subjective experience of a unitary nature in our perceptual world and seemingly discrete actions, most careful observations of behavioral decisions and even their associated neural encoding point, in many cases, to a more probabilistic view (Averbeck et al., 2006). 'Noise' often becomes an important component of such systems, for instance in gain encoding (Knill & Pouget, 2004). Neuronal noise is often near-Poisson (Tolhurst et al., 1983), and while it would seem that the tremendous variability often seen in neuronal firing is detrimental, models suggest that Poisson noise allows a population code to represent both the mean and variance of the variables being encoded - something crucial for Bayesian inference.

While these results are from behavior and ensembles of neurons, the concepts would very possibly scale to those at the whole brain. Probably the most closely related work at that scale is by Karl Friston (K. Friston, 2012), where Bayesian Variational methods are used as an explanatory mechanism for embodied perception. Recently this approach was used as a theoretical basis for resting state activity (K. J. Friston et al., 2014). They consider a system undergoing self-organized instability and critical slowing. This assumes that a system is minimising the entropy of exogenous fluctuations driving internal states through minimising self-information, a kind of 'surprise' in the system, defined through a free energy function quantifying this as the error in internal representations compared to sensory input. Briefly, this model shows that an expression of free energy defines a kind of landscape where minimums never have a high curvature. That is, internal states will tend to flow into these valleys, but are not constrained by free energy and can be sensitive to random fluctuations. They suggest that from a Bayesian perspective this self-

organized instability reflects that a system seeks to respond sensitively to sensory pertubations while resisting an overly precise or particular interpretation. This can be seen as a mechanism for metastability and allows for a more flexible repertoire of action (Breakspear, 2004).

Conclusion

Tremendous challenges remain in bettering our understanding of the relationship between the brain, perception, and action. The rapid advancement of technologies used to record and image the brain has resulted in the need for increasingly complex models of that data. Recently the BRAIN initiative has called for an acceleration in developing these recording technologies, as well as investing in creating the theoretical, computational, and machine learning tools and knowledge that will be necessary to handle new data. One might imagine the ability to look into a brain with perfect electrical knowledge, which somehow underlies the full spectrum of cognition and behavior. With our current methods and prior knowledge this data would be completely overwhelming. These developments parallel those outside academia, where "big data" has become a new industry, and many companies seek to mine ever increasing amounts of data for useful, and profitable, information.

As the models we use to leverage our understanding of data collected from neural systems become more complex there may be a tendency to create "black box" systems that are successful at predicting outcome, but in a mysterious fashion as the parameters are too complicated to easily interpret. This may be appropriate for some cases, such as most industry applications, or perhaps in diagnostic tools, but basic science questions require methods that can be disentangled to under-

stand some mechanistic basis. The brain may be a special case, as its a peculiar situation where we are developing artificial learning tools to understand what is essentially a biological "black box", and if these tools add to that complexity we may be heading in the wrong direction.

We are rapidly moving into a period of "big data" neuroscience, and it may be a period marked by every answer leading to a new set of questions. Not long ago spontaneous and resting activity in the brain was viewed largely as noise to average away, but it was discovered that it was actually well organized and not very different from the task-oriented brain. More than a decade after the term "default mode" was coined, this work adds to a small but growing body of evidence that there is yet another level of complexity to be examined in spontaneous activity, and this leads to numerous questions in itself. These questions will hopefully help identify new avenues to explore with future experiments, and better prepare us for the time when we do have a more complete window into the brain.

List of References

Achard, S., Salvador, R., Whitcher, B., Suckling, J., & Bullmore, E. (2006, January). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *26*(1), 63–72.

Alexander, G. E., & Crutcher, M. D. (1990, July). Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *Trends in neurosciences*, *13*(7), 266–71.

Allen, E. a., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., & Calhoun, V. D. (2014, March). Tracking whole-brain connectivity dynamics in the resting state. *Cerebral cortex (New York, N.Y. : 1991)*, *24*(3), 663–76.

Anderson, J. S., Nielsen, J. A., Froehlich, A. L., DuBray, M. B., Druzgal, T. J., Cariello, A. N., ... Lainhart, J. E. (2011, December). Functional connectivity magnetic resonance imaging classification of autism. *Brain : a journal of neurology*, *134*(Pt 12), 3742–54.

Andrew Y. Ng, Y. W., Michael I. Jordan. (n.d.). On Spectral Clustering: Analysis and an algorithm.

Arfanakis, K., Cordes, D., Haughton, V. M., Moritz, C. H., Quigley, M. A., & Meyerand, M. E. (2000, October). Combining independent component analysis and correlation analysis to probe interregional connectivity in fMRI task activation datasets. *Magnetic resonance imaging*, *18*(8), 921–30.

Averbeck, B. B., Latham, P. E., & Pouget, A. (2006, May). Neural correlations, population coding and computation. *Nature reviews. Neuroscience*, *7*(5), 358–66.

Baltes, P. B., & Lindenberger, U. (n.d.). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging?

Barabasi, A. (1999). Emergence of Scaling in Random Networks. *Science*, *286*(October), 509–512.

Barbieri, D., Citti, G., Cocci, G., & Sarti, A. (2014, January). A Cortical-Inspired Geometry for Contour Perception and Motion Integration. *Journal of Mathematical Imaging and Vision*, *49*(3), 511–529.

Bassett, D. S., Wymbs, N. F., Porter, M. A., Mucha, P. J., Carlson, J. M., & Grafton, S. T. (2011, May). Dynamic reconfiguration of human brain networks during learning. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(18), 7641–6.

Baum, L. (1972). *An Inequality and Associated Maximization Technique in Statistical Estimation of Probabilistic Functions of a Markov Process.*

Baum, L., & Petrie, T. (1966). *JSTOR: The Annals of Mathematical Statistics, Vol. 37, No. 6 (Dec., 1966), pp. 1554-1563.*

Belmonte, M. K., Allen, G., Beckel-Mitchener, A., Boulanger, L. M., Carper, R. a., & Webb, S. J. (2004, October). Autism and abnormal development of brain connectivity. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *24*(42), 9228–31.

Bishop, C. M. (2007). *Pattern Recognition and Machine Learning.*

Boots, B., Siddiqi, S. M., & Gordon, G. J. (2009, December). Closing the Learning-Planning Loop with Predictive State Representations.

Boots, B., Siddiqi, S. M., & Gordon, G. J. (2011, June). Closing the learning-planning loop with predictive state representations. *The International Journal of Robotics Research*, *30*(7), 954–966.

Breakspear, M. (2004, January). "Dynamic" connectivity in neural systems: theoretical and empirical considerations. *Neuroinformatics*, *2*(2), 205–26.

Bressler, S. L., Tang, W., Sylvester, C. M., Shulman, G. L., & Corbetta, M. (2008, October). Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *28*(40), 10056–61.

Broyd, S. J., Demanuele, C., Debener, S., Helps, S. K., James, C. J., & Sonuga-Barke, E. J. S. (2009, March). Default-mode brain dysfunction in mental disorders: a systematic review. *Neuroscience and biobehavioral reviews*, *33*(3), 279–96.

Bullmore, E., & Sporns, O. (2009, March). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews. Neuroscience*, *10*(3), 186–98.

Chai, X. J., Castañón, A. N., Ongür, D., & Whitfield-Gabrieli, S. (2012, January). Anticorrelations in resting state networks without global signal regression. *NeuroImage*, *59*(2), 1420–8.

Chang, C., & Glover, G. H. (2010, March). Time-frequency dynamics of resting-state brain connectivity measured with fMRI. *NeuroImage*, *50*(1), 81–98.

Chen, G., Chen, G., Xie, C., Ward, B. D., Li, W., Antuono, P., & Li, S.-J. (2012, December). A method to determine the necessity for global signal regression in resting-state fMRI studies. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *68*(6), 1828–35.

Cheng, K., Waggoner, R. A., & Tanaka, K. (2001, October). Human ocular dominance columns as revealed by high-field functional magnetic resonance imaging. *Neuron*, *32*(2), 359–74.

Chung, F. R. K. (1994). Lectures on Spectral Graph Theory. *CBMS Regional Conference Series in Mathematics, No. 92*.

Citti, G., & Sarti, A. (2014). *Neuromathematics of Vision.*

Cohen Kadosh, K., Henson, R. N. A., Cohen Kadosh, R., Johnson, M. H., & Dick, F. (2010, May). Task-dependent activation of face-sensitive cortex: an fMRI adaptation study. *Journal of cognitive neuroscience*, *22*(5), 903–17.

Cohen Kadosh, K., & Johnson, M. H. (2007, September). Developing a cortex specialized for face perception. *Trends in cognitive sciences*, *11*(9), 367–9.

Coifman, R. R., Lafon, S., Lee, a. B., Maggioni, M., Nadler, B., Warner, F., & Zucker, S. W. (2005, May). Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(21), 7426–31.

Collins, H. R., Zhu, X., Bhatt, R. S., Clark, J. D., & Joseph, J. E. (2012). Process and Domain Specificity in RegionsEngaged for Face Processing: An fMRI Study of Perceptual Differentiation. *Journal of Cognitive Neuroscience*, 1–17.

Cordes, D., Haughton, V. M., Arfanakis, K., Carew, J. D., Turski, P. A., Moritz, C. H., ... Meyerand, M. E. (2001, August). Frequencies Contributing to Functional Connectivity in the Cerebral Cortex in "Resting-state" Data. *AJNR Am. J. Neuroradiol.*, *22*(7), 1326–1333.

Cordes, D., Haughton, V. M., Arfanakis, K., Wendt, G. J., Turski, P. a., Moritz, C. H., ... Meyerand, M. E. (2000, October). Mapping functionally related regions of brain with functional connectivity MR imaging. *AJNR. American journal of neuroradiology*, *21*(9), 1636–44.

Cribben, I., Haraldsdottir, R., Atlas, L. Y., Wager, T. D., & Lindquist, M. a. (2012, July). Dynamic connectivity regression: determining state-related changes in brain connectivity. *NeuroImage*, *61*(4), 907–20.

Damoiseaux, J. S., Rombouts, S. A. R. B., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., & Beckmann, C. F. (2006, September). Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(37), 13848–53.

Danóczy, M., & Hahnloser, R. H. R. (2005, January). Efficient estimation of hidden state dynamics from spike trains. In *Advances in neural information processing systems 18 [neural information processing systems, nips 2005, december 5-8, 2005, vancouver, british columbia, canada]*.

De Luca, M., Beckmann, C. F., De Stefano, N., Matthews, P. M., & Smith, S. M. (2006, February). fMRI resting state networks define distinct modes of long-distance interactions in the human brain. *NeuroImage*, *29*(4), 1359–67.

Deco, G., & Jirsa, V. K. (2012, March). Ongoing cortical activity at rest: criticality, multistability, and ghost attractors. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *32*(10), 3366–75.

Deco, G., Jirsa, V. K., & McIntosh, A. R. (2011, January). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature reviews. Neuroscience*, *12*(1), 43–56.

Deco, G., Jirsa, V. K., & McIntosh, A. R. (2013, May). Resting brains never rest: computational insights into potential cognitive architectures. *Trends in neurosciences*, *36*(5), 268–74.

Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Wilson, J., Moss, H. E., ... Tyler, L. K. (2000, June). Susceptibility-induced loss of signal: comparing PET and fMRI on a semantic task. *NeuroImage*, *11*(6 Pt 1), 589–600.

Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., ... Milham, M. P. (2014, June). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, *19*(6), 659–67.

Dosenbach, N. U. F., Nardos, B., Cohen, A. L., Fair, D. a., Power, J. D., Church, J. a., ... Schlaggar, B. L. (2010, September). Prediction of individual brain maturity using fMRI. *Science (New York, N.Y.)*, *329*(5997), 1358–61.

Dougherty, R. F., Koch, V. M., Brewer, A. A., Fischer, B., Modersitzki, J., & Wandell, B. A. (2003, January). Visual field representations and locations of visual areas V1/2/3 in human visual cortex. *Journal of vision*, *3*(10), 586–98.

Doya, K. (1999, October). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural networks : the official journal of the International Neural Network Society*, *12*(7-8), 961–974.

Eavani, H., Satterthwaite, T. D., Gur, R. E., Gur, R. C., & Davatzikos, C. (2013, January). Unsupervised learning of functional network dynamics in resting state fMRI. *Information processing in medical imaging : proceedings of the ... conference*, *23*, 426–37.

Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank.

Ecker, C., Rocha-Rego, V., Johnston, P., Mourao-Miranda, J., Marquand, A., Daly, E. M., ... Murphy, D. G. (2010, January). Investigating the predictive value of whole-brain structural MR scans in autism: a pattern classification approach. *NeuroImage*, *49*(1), 44–56.

Ernst Mach, C. M. W. (1897). *Contributions To The Analysis Of The Sensations.*

Estivill-Castro, V. (2002, June). Why so many clustering algorithms. *ACM SIGKDD Explorations Newsletter*, *4*(1), 65–75.

Fair, D. a., Cohen, A. L., Dosenbach, N. U. F., Church, J. a., Miezin, F. M., Barch, D. M., ... Schlaggar, B. L. (2008, March). The maturing architecture of the brain's default network. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(10), 4028–32.

Fair, D. a., Cohen, A. L., Power, J. D., Dosenbach, N. U. F., Church, J. a., Miezin, F. M., ... Petersen, S. E. (2009, May). Functional brain networks develop from a "local to distributed" organization. *PLoS computational biology*, *5*(5), e1000381.

Fair, D. a., Dosenbach, N. U. F., Church, J. a., Cohen, A. L., Brahmbhatt, S., Miezin, F. M., ... Schlaggar, B. L. (2007, August). Development of distinct control networks through segregation and integration. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(33), 13507–12.

Faisan, S., Thoraval, L., Armspach, J.-P., & Heitz, F. (2007, February). Hidden Markov multiple event sequence models: A paradigm for the spatio-temporal analysis of fMRI data. *Medical image analysis*, *11*(1), 1–20.

Figley, C. R., & Stroman, P. W. (2011, February). The role(s) of astrocytes and astrocyte activity in neurometabolism, neurovascular coupling, and the production of functional neuroimaging signals. *The European journal of neuroscience*, *33*(4), 577–88.

Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L., & Raichle, M. E. (2006, June). Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(26), 10046–51.

Fox, M. D., & Raichle, M. E. (2007, September). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature reviews. Neuroscience*, *8*(9), 700–11.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005, July). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(27), 9673–8.

Fox, M. D., Snyder, A. Z., Vincent, J. L., & Raichle, M. E. (2007, October). Intrinsic fluctuations within cortical systems account for intertrial variability in human behavior. *Neuron*, *56*(1), 171–84.

Fox, M. D., Snyder, A. Z., Zacks, J. M., & Raichle, M. E. (2006, January). Coherent spontaneous activity accounts for trial-to-trial variability in human evoked brain responses. *Nature neuroscience*, *9*(1), 23–5.

Fox, M. D., Zhang, D., Snyder, A. Z., & Raichle, M. E. (2009, June). The global signal and observed anticorrelated resting state brain networks. *Journal of neurophysiology*, *101*(6), 3270–83.

Frank, M. J., & Claus, E. D. (2006, April). Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychological review*, *113*(2), 300–26.

Friston, K. (2012, November). A Free Energy Principle for Biological Systems. *Entropy (Basel, Switzerland)*, *14*(11), 2100–2121.

Friston, K., Kilner, J., & Harrison, L. (2006, January). A free energy principle for the brain. *Journal of physiology, Paris*, *100*(1-3), 70–87.

Friston, K. J., Harrison, L., & Penny, W. (2003, August). Dynamic causal modelling. *NeuroImage*, *19*(4), 1273–302.

Friston, K. J., Kahan, J., Razi, A., Stephan, K. E., & Sporns, O. (2014, May). On nodes and modes in resting state fMRI. *NeuroImage*, *99*, 533–47.

Garrett, D. D., Kovacevic, N., McIntosh, A. R., & Grady, C. L. (2010, April). Blood oxygen level-dependent signal variability is more than just noise. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *30*(14), 4914–21.

Garrett, D. D., Kovacevic, N., McIntosh, A. R., & Grady, C. L. (2011, March). The importance of being variable. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *31*(12), 4496–503.

Gephi: an open source software for exploring and manipulating networks. (n.d.).

Ginestet, C. E., & Simmons, A. (2011, March). Statistical parametric network analysis of functional connectivity dynamics during a working memory task. *NeuroImage*, *55*(2), 688–704.

Golland, Y., Golland, P., Bentin, S., & Malach, R. (2008, January). Data-driven clustering reveals a fundamental subdivision of the human cortex into two global systems. *Neuropsychologia*, *46*(2), 540–53.

Goodale, M. A., Milner, A. D., Jakobson, L. S., & Carey, D. P. (1991, January). A neurological dissociation between perceiving objects and grasping them. *Nature*, *349*(6305), 154–6.

Gour, N., Ranjeva, J.-P., Ceccaldi, M., Confort-Gouny, S., Barbeau, E., Soulier, E., ... Felician, O. (2011, September). Basal functional connectivity within the anterior temporal network is associated with performance on declarative memory tasks. *NeuroImage*, *58*(2), 687–97.

Greicius, M. D., Flores, B. H., Menon, V., Glover, G. H., Solvason, H. B., Kenna, H., ... Schatzberg, A. F. (2007, September). Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biological psychiatry*, *62*(5), 429–37.

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003, January). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(1), 253–8.

Greicius, M. D., Srivastava, G., Reiss, A. L., & Menon, V. (2004, March). Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(13), 4637–42.

Hasselmo, M. E., Rolls, E. T., & Baylis, G. C. (1989, April). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioural brain research*, *32*(3), 203–18.

Haxby, J., Hoffman, E., & Gobbini, M. (2000, June). The distributed human neural system for face perception. *Trends in cognitive sciences*, *4*(6), 223–233.

Helmholtz, H. L. (1910). *Helmholtz'a Treatise on Physiological Optics, Volume III: The Perceptions of Vision.*

Herbst, J. A., Gammeter, S., Ferrero, D., & Hahnloser, R. H. R. (2008, September). Spike sorting with hidden Markov models. *Journal of neuroscience methods*, *174*(1), 126–34.

Honey, C. J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J. P., Meuli, R., & Hagmann, P. (2009, February). Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(6), 2035–40.

Honey, C. J., Thivierge, J.-P., & Sporns, O. (2010, September). Can structure predict function in the human brain? *NeuroImage*, *52*(3), 766–76.

Hsu, D., Kakade, S. M., & Zhang, T. (2012, September). A spectral algorithm for learning Hidden Markov Models. *Journal of Computer and System Sciences*, *78*(5), 1460–1480.

Hubel, D. H., & Wiesel, T. N. (1959, October). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, *148*, 574–91.

Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., ... Chang, C. (2013, October). Dynamic functional connectivity: promise, issues, and interpretations. *NeuroImage*, *80*, 360–78.

Jaeger, H. (2000, June). Observable operator models for discrete stochastic time series. *Neural computation*, *12*(6), 1371–98.

Jones, D. T., Vemuri, P., Murphy, M. C., Gunter, J. L., Senjem, M. L., Machulda, M. M., ... Jack, C. R. (2012, January). Non-stationarity in the "resting brain's" modular architecture. *PloS one*, *7*(6), e39731.

Just, M. A., Cherkassky, V. L., Keller, T. A., & Minshew, N. J. (2004, August). Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity. *Brain : a journal of neurology*, *127*(Pt 8), 1811–21.

Kalivas, P. W., & Volkow, N. D. (2005, August). The neural basis of addiction: a pathology of motivation and choice. *The American journal of psychiatry*, *162*(8), 1403–13.

Kalman, R. E. (1960, March). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, *82*(1), 35.

Kelly, a. M. C., Uddin, L. Q., Biswal, B. B., Castellanos, F. X., & Milham, M. P. (2008, January). Competition between functional brain networks mediates behavioral variability. *NeuroImage*, *39*(1), 527–37.

Kitazawa, S., Kimura, T., & Yin, P. B. (1998, April). Cerebellar complex spikes encode both destinations and errors in arm movements. *Nature*, *392*(6675), 494–7.

Knill, D. C., & Pouget, A. (2004, December). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in neurosciences*, *27*(12), 712–9.

Koch, M. A., Norris, D. G., & Hund-Georgiadis, M. (2002, May). An investigation of functional and anatomical connectivity using magnetic resonance imaging. *NeuroImage*, *16*(1), 241–50.

Kucyi, A., & Davis, K. D. (2014, October). Dynamic functional connectivity of the default mode network tracks daydreaming. *NeuroImage*, *100*, 471–80.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., ... Turner, R. (1992, June). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences*, *89*(12), 5675–5679.

Lange, N., Dubray, M. B., Lee, J. E., Froimowitz, M. P., Froehlich, A., Adluru, N., ... Lainhart, J. E. (2010, December). Atypical diffusion tensor hemispheric asymmetry in autism. *Autism research : official journal of the International Society for Autism Research*, *3*(6), 350–8.

Laskaris, N. A., Liu, L. C., & Ioannides, A. A. (2003, October). Single-trial variability in early visual neuromagnetic responses: an explorative study based on the regional activation contributing to the N70m peak. *NeuroImage*, *20*(2), 765–83.

Ledoit, O., & Wolf, M. (2003, December). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, *10*(5), 603–621. doi: 10.1016/S0927-5398(03)00007-0

Lee, J. H., Durand, R., Gradinaru, V., Zhang, F., Goshen, I., Kim, D.-S., ... Deisseroth, K. (2010, June). Global and local fMRI signals driven by neurons defined optogenetically by type and wiring. *Nature*, *465*(7299), 788–92.

Leo, A., Bernardi, G., Handjaras, G., Bonino, D., Ricciardi, E., & Pietrini, P. (2012, January). Increased BOLD variability in the parietal cortex and enhanced parieto-occipital connectivity during tactile perception in congenitally blind individuals. *Neural plasticity*, *2012*, 720278.

Leonardi, N., Richiardi, J., Gschwind, M., Simioni, S., Annoni, J.-M., Schluep, M., ... Van De Ville, D. (2013, December). Principal components of functional connectivity: a new approach to study dynamic brain connectivity during rest. *NeuroImage*, *83*, 937–50.

Littman, M. L., & Sutton, R. S. (2002). Predictive Representations of State. In *Advances in neural information processing systems* (pp. 1555–1561).

Littman, M. L., Sutton, R. S., & Singh, S. (2001). Predictive Representatins of State. *NIPS*, *14*, 1555–1561.

Logothetis, N. K. (2003, May). The Underpinnings of the BOLD Functional Magnetic Resonance Imaging Signal. *J. Neurosci.*, *23*(10), 3963–3971.

Logothetis, N. K. (2010, November). Bold claims for optogenetics. *Nature*, *468*(7323), E3–4; discussion E4–5.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001, July). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*(6843), 150–7.

Logothetis, N. K., & Wandell, B. a. (2004, January). Interpreting the BOLD signal. *Annual review of physiology*, *66*, 735–69.

Luxburg, U. (2007a, August). A tutorial on spectral clustering. *Statistics and Computing*, *17*(4), 395–416.

Luxburg, U. (2007b, August). A tutorial on spectral clustering. *Statistics and Computing*, *17*(4), 395–416.

Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003, May). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron*, *38*(3), 487–97.

Marr, D. (1969, June). A theory of cerebellar cortex. *The Journal of physiology*, *202*(2), 437–70.

Marrelec, G., Krainik, A., Duffau, H., Pélégrini-Issac, M., Lehéricy, S., Doyon, J., & Benali, H. (2006, August). Partial correlation for functional brain interactivity investigation in functional MRI. *NeuroImage*, *32*(1), 228–37.

Martin, A., & Chao, L. L. (2001, April). Semantic memory and the brain: structure and processes. *Current opinion in neurobiology*, *11*(2), 194–201.

McIntosh, A. R., Kovacevic, N., & Itier, R. J. (2008, January). Increased brain signal variability accompanies lower behavioral variability in development. *PLoS computational biology*, *4*(7), e1000106.

McIntosh, R. D., & Schenk, T. (2009, May). Two visual streams for perception and action: current trends. *Neuropsychologia*, *47*(6), 1391–6.

Mehrkanoon, S., Breakspear, M., & Boonstra, T. W. (2014, May). Low-dimensional dynamics of resting-state cortical activity. *Brain topography*, *27*(3), 338–52.

Milner, D. (n.d.). *The Visual Brain in Action: Paperback: David Milner - Oxford University Press.*

Mohr, P. N. C., & Nagel, I. E. (2010, June). Variability in brain activity as an individual difference measure in neuroscience? *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *30*(23), 7755–7.

Moser, E. I., Kropff, E., & Moser, M.-B. (2008, January). Place cells, grid cells, and the brain's spatial representation system. *Annual review of neuroscience*, *31*, 69–89.

Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I., & Malach, R. (2005, August). Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. *Science (New York, N.Y.)*, *309*(5736), 951–4.

Ng, Jordan, W. (2002). On Spectral Clustering.

Nielsen, J. A., Zielinski, B. A., Fletcher, P. T., Alexander, A. L., Lange, N., Bigler, E. D., ... Anderson, J. S. (2013, January). Multisite functional connectivity MRI classification of autism: ABIDE results. *Frontiers in human neuroscience*, *7*, 599.

Nir, Y., Mukamel, R., Dinstein, I., Privman, E., Harel, M., Fisch, L., ... Malach, R. (2008, August). Interhemispheric correlations of slow spontaneous neuronal fluctuations revealed in human sensory cortex. *Nature Neuroscience*, *11*(9), 1100–1108.

Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., & Ugurbil, K. (1992, July). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(13), 5951–5.

O'Keefe, J., & Dostrovsky, J. (1971, November). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, *34*(1), 171–175.

Parga, N., & Abbott, L. F. (2007, November). Network model of spontaneous activity exhibiting synchronous transitions between up and down States. *Frontiers in neuroscience*, *1*(1), 57–66.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007, December). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature reviews. Neuroscience*, *8*(12), 976–87.

Perrett, D. I., Rolls, E. T., & Caan, W. (1982, January). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *47*(3), 329–42.

Pobric, G., Jefferies, E., & Ralph, M. a. L. (2007, December). Anterior temporal lobes mediate semantic representation: mimicking semantic dementia by using rTMS in normal participants. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(50), 20137–41.

Pouget, A., Dayan, P., & Zemel, R. S. (2003, January). Inference and computation with population codes. *Annual review of neuroscience*, *26*, 381–410.

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012, February). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, *59*(3), 2142–54. doi: 10.1016/j.neuroimage.2011.10.018

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., ... Petersen, S. E. (2011, November). Functional network organization of the human brain. *Neuron*, *72*(4), 665–78.

Quiroga, R. Q., Kreiman, G., Koch, C., & Fried, I. (2008, March). Sparse but not 'grandmother-cell' coding in the medial temporal lobe. *Trends in cognitive sciences*, *12*(3), 87–91.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005, June). Invariant visual representation by single neurons in the human brain. *Nature*, *435*(7045), 1102–7.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Raichle, M. E. (2009, October). A paradigm shift in functional brain imaging. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *29*(41), 12729–34.

Raichle, M. E., MacLeod, a. M., Snyder, a. Z., Powers, W. J., Gusnard, D. a., & Shulman, G. L. (2001, January). A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(2), 676–82.

Raj, A., Kuceyeski, A., & Weiner, M. (2012, March). A network diffusion model of disease progression in dementia. *Neuron*, *73*(6), 1204–15.

Roberts, T. P. L., Khan, S. Y., Rey, M., Monroe, J. F., Cannon, K., Blaskey, L., ... Edgar, J. C. (2010, February). MEG detection of delayed auditory evoked responses in autism spectrum disorders: towards an imaging biomarker for autism. *Autism research : official journal of the International Society for Autism Research*, *3*(1), 8–18.

Roebroeck, A., Formisano, E., & Goebel, R. (2005, March). Mapping directed influence over the brain using Granger causality and fMRI. *NeuroImage*, *25*(1), 230–42.

Rolls, E. T. (2013, January). The mechanisms for pattern completion and pattern separation in the hippocampus. *Frontiers in systems neuroscience*, *7*, 74.

Rosencrantz, M., Gordon, G., & Thrun, S. (2004, July). Learning low dimensional predictive representations. In *Twenty-first international conference on machine learning - icml '04* (p. 88). New York, New York, USA: ACM Press.

Rubinov, M., & Sporns, O. (2010, September). Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*, *52*(3), 1059–69.

Rudie, J. D., Shehzad, Z., Hernandez, L. M., Colich, N. L., Bookheimer, S. Y., Iacoboni, M., & Dapretto, M. (2012, May). Reduced functional integration and segregation of distributed neural systems underlying social and emotional information processing in autism spectrum disorders. *Cerebral cortex (New York, N.Y. : 1991)*, *22*(5), 1025–37.

Salvador, R., Suckling, J., Coleman, M. R., Pickard, J. D., Menon, D., & Bullmore, E. (2005, September). Neurophysiological architecture of functional magnetic resonance images of human brain. *Cerebral cortex (New York, N.Y. : 1991)*, *15*(9), 1332–42.

Samanez-Larkin, G. R., Kuhnen, C. M., Yoo, D. J., & Knutson, B. (2010, January). Variability in nucleus accumbens activity mediates age-related suboptimal financial risk taking. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *30*(4), 1426–34.

Sch, J. (2005). Statistical Applications in Genetics and Molecular Biology A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Func.

Schneider, K. A., Richter, M. C., & Kastner, S. (2004, October). Retinotopic organization and functional subdivisions of the human lateral geniculate nucleus: a high-resolution functional magnetic resonance imaging study. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *24*(41), 8975–85.

Schultz, W. (1998, July). Predictive Reward Signal of Dopamine Neurons. *J Neurophysiol*, *80*(1), 1–27.

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., ... Greicius, M. D. (2007, February). Dissociable intrinsic connectivity networks for salience processing and executive control. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *27*(9), 2349–56.

Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004, March). Inversion leads to quantitative, not qualitative, changes in face processing. *Current biology : CB*, *14*(5), 391–6.

Shulman, G. L., Fiez, J. a., Corbetta, M., Buckner, R. L., Miezin, F. M., Raichle, M. E., & Petersen, S. E. (1997, October). Common Blood Flow Changes across Visual Tasks: II. Decreases in Cerebral Cortex. *Journal of Cognitive Neuroscience*, *9*(5), 648–663.

Siddiqi, S. M., Boots, B., & Gordon, G. J. (2009, October). Reduced-Rank Hidden Markov Models.

Singh, S., James, M. R., & Rudary, M. R. (2004, July). Predictive state representations: a new theory for modeling dynamical systems. , 512–519.

Smith, S. M. (2012, August). The future of FMRI connectivity. *NeuroImage*, *62*(2), 1257–66.

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., ... Woolrich, M. W. (2011, January). Network modelling methods for FMRI. *NeuroImage*, *54*(2), 875–91.

Stigler, J., Ziegler, F., Gieseke, A., Gebhardt, J. C. M., & Rief, M. (2011, October). The complex folding network of single calmodulin molecules. *Science (New York, N.Y.)*, *334*(6055), 512–6.

Sutton, R. S. (1998). *Sutton & Barto Book: Reinforcement Learning: An Introduction.*

Sylvester, R., Josephs, O., Driver, J., & Rees, G. (2007, February). Visual FMRI responses in human superior colliculus show a temporal-nasal asymmetry that is absent in lateral geniculate and visual cortex. *Journal of neurophysiology*, *97*(2), 1495–502.

Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007, December). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, *35*(6), 2769–2794. Retrieved from http://projecteuclid.org/euclid.aos/1201012979

Thompson, G. J., Merritt, M. D., Pan, W.-J., Magnuson, M. E., Grooms, J. K., Jaeger, D., & Keilholz, S. D. (2013, December). Neural correlates of time-varying functional connectivity in the rat. *NeuroImage*, *83*, 826–36.

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983, January). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision research*, *23*(8), 775–85.

Uddin, L. Q., Kelly, a. M., Biswal, B. B., Xavier Castellanos, F., & Milham, M. P. (2009, February). Functional connectivity of default mode network components: correlation, anticorrelation, and causality. *Human brain mapping*, *30*(2), 625–37.

Vempala, S., & Wang, G. (2004, June). A spectral algorithm for learning mixture models. *Journal of Computer and System Sciences*, *68*(4), 841–860.

Ventura-Campos, N., Sanjuán, A., González, J., Palomar-García, M.-A., Rodríguez-Pujadas, A., Sebastián-Gallés, N., ... Ávila, C. (2013, May). Spontaneous brain activity predicts learning ability of foreign sounds. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *33*(22), 9295–305.

Vigneau-Roy, N., Bernier, M., Descoteaux, M., & Whittingstall, K. (2014, May). Regional variations in vascular density correlate with resting-state and task-evoked blood oxygen level-dependent signal amplitude. *Human brain mapping*, *35*(5), 1906–20.

Vincent, J. L., Patel, G. H., Fox, M. D., Snyder, a. Z., Baker, J. T., Van Essen, D. C., ... Raichle, M. E. (2007, May). Intrinsic functional architecture in the anaesthetized monkey brain. *Nature*, *447*(7140), 83–6.

Visser, M., Jefferies, E., & Lambon Ralph, M. a. (2010, June). Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. *Journal of cognitive neuroscience*, *22*(6), 1083–94.

Viswanathan, A., & Freeman, R. D. (2007, October). Neurometabolic coupling in cerebral cortex reflects synaptic more than spiking activity. *Nature neuroscience*, *10*(10), 1308–12.

Vossel, S., Weidner, R., Driver, J., Friston, K. J., & Fink, G. R. (2012, August). Deconstructing the architecture of dorsal and ventral attention systems with dynamic causal modeling. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *32*(31), 10637–48.

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007, October). Visual field maps in human cortex. *Neuron*, *56*(2), 366–83.

Wang, H., Chen, C., & Fushing, H. (2012, January). Extracting multiscale pattern information of fMRI based functional brain connectivity with application on classification of autism spectrum disorders. *PloS one*, *7*(10), e45502.

Welchew, D. E., Ashwin, C., Berkouk, K., Salvador, R., Suckling, J., Baron-Cohen, S., & Bullmore, E. (2005, May). Functional disconnectivity of the medial temporal lobe in Asperger's syndrome. *Biological psychiatry*, *57*(9), 991–8.

Wong, C., & Gallate, J. (2012, April). The function of the anterior temporal lobe: a review of the empirical evidence. *Brain research*, *1449*, 94–116.

Xia, M., Wang, J., & He, Y. (2013, January). BrainNet Viewer: a network visualization tool for human brain connectomics. *PloS one*, *8*(7), e68910.

Yacoub, E., Harel, N., & Ugurbil, K. (2008, July). High-field fMRI unveils orientation columns in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(30), 10607–12.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of experimental psychology*, *81*(1), 141–145.

Zalesky, A., Fornito, A., Cocchi, L., Gollo, L. L., & Breakspear, M. (2014, June). Time-resolved resting-state brain networks. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(28), 10341–6.

Zhao, H., & Poupart, P. (2014, June). A Sober Look at Spectral Learning. *ArXiv*([cs.LG]), 1406.4631.